WZ-SPEED HARMONIZER: AN OPTIMIZED ACTIVE TRAFFIC AND DEMAND MANAGEMENT SYSTEM WITH SPEED HARMONIZATION FOR WORK ZONES

BY

HANI RAMEZANI

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Civil Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2016

Urbana, Illinois

Doctoral Committee:

Professor Rahim F. Benekohal, Chair
Associate Professor Yanfeng Ouyang
Assistant Professor Daniel B. Work
Associate Professor Angelia Nedich

Abstract

Speed harmonization, known as Variable Speed Limit (VSL), is implemented through a number of Changeable Message Signs (CMSs) spaced out over a stretch of highway. The CMSs display advisory speeds that may change over time and space to regulate travel speed and arrival time to a highway bottleneck and consequently reduce congestion impacts. Usually speed harmonization studies determine optimal dynamic advisory speeds to minimize travel time and delay. Delay and travel time can be further minimized if optimal location and number of CMSs are determined as well. Determining dynamic advisory speeds depends on the number and locations of CMSs; thus these variables have to be optimized simultaneously. This study is the first study which simultaneously optimized 1) dynamic advisory speeds, 2) number of CMSs, and 3) location of the CMSs to minimize total travel time and a penalty function for the number of CMSs. This problem is important as it reduces maintenance and installation costs of CMSs and enhances effectiveness of speed harmonization to reduce congestion impacts such as delay. Solving this problem is challenging because it is a large scale Mixed Integer Nonlinear Program (MINLP).

Determining traffic speed and extend of queue is critical to develop a speed harmonization scheme to mitigate congestion. Past studies have used first order model or second order model as a constraint in the optimization program to determine speed and extend of queue. This study compared the first order model and the second order model versus field data and showed that the second order model was better in estimating average queue length and maximum queue length.

To use the second order mode, it is necessary to calibrate the model parameters, which are relaxation time (τ) and anticipation coefficient (ϑ). This study calibrated the parameters using work zone field data to minimize error in speed and queue length estimates. The calibration is a complex process since there might be many local optimal points returning parameter values that are not physically justifiable. To overcome this issue, this study proposed a new calibration procedure. The methodology detected the behavior of the second order model in the τ- ϑ space and determined a search direction and its boundaries to avoid stopping at local minima.

Although the second order model returned acceptable average and maximum queue lengths, it returned slower queue propagation pattern and faster queue dissipation pattern than

field data. Thus the two- ϑ model was proposed to more maturely reflect asymmetric queue propagation and dissipation. The modification considered two different anticipation coefficients for queue propagation (ϑ_p) and queue shrinkage (ϑ_s). The two- ϑ model was calibrated using field data and results showed that the ratio of $\frac{\vartheta_s}{\vartheta_p}$ ranges from 1.86 to 2.6 and both of them are greater than single ϑ.

The solutions for the optimization program mainly tackled integrality. One source of integrality is piecewise speed-density models. Previous studies have used single-regime (single piece) models, but these models are not generally sufficient to describe congested conditions. Thus this study included the piecewise models as constraints to enhance accuracy of traffic state prediction in congested conditions. A continuous transformation approach was proposed to eliminate relevant integer variables. The methodology was used to optimize speed harmonization for a 12.5-mile roadway and a 50-min analysis duration where locations of the CMSs were given. For this problem, the methodology eliminated 30,000 binary variables and solved the problem in 23 minutes when the program included roughly 158,000 variables and 186,000 constraints. The results showed that WZSH can reduce delay by 15.7% and maximum queue length by 37.5% compared to the no speed harmonization condition.

Another source of integrality is binary variables to determine location and number of CMSs. To solve the problem, three solution methods were proposed: 1) Greedy algorithm, 2) Augmented-Cut and Branch (AC&B) method, and 3) Approximate Decomposition method. The three solution methods were compared using a benchmark problem and results showed that the three methods return very close objective function values (within 1%), but the Approximate Decomposition method requires less computational resources. In particular, the Approximate Decomposition method reduced the number of WZSH solution by factors of 14.1 and 6.6 compared with Greedy Algorithm and AC&B method, respectively.

Table of Contents

Chapter 1 Introduction

Speed harmonization, known as Variable Speed Limit (VSL), is implemented through a number of Changeable Message Signs (CMSs) spaced out over a stretch of highway. The CMSs display advisory speeds that may change over time and space to regulate travel speed and improve traffic operation and safety. The Federal Highway Administration (FHWA) introduced the Active Traffic and Demand Management (ATDM) strategies to manage traffic congestion. According to FHWA, "ATDM is the dynamic management, control, and influence of travel demand, traffic demand, and traffic flow of transportation facilities." (Federal Highway Administration, 2012). An example of the ATDM strategies is speed harmonization which is the focus area of this study.

Speed harmonization can improve traffic operation by increasing capacity or reducing the shockwave propagation rate. Capacity may be increased when speed harmonization reduces speed difference between vehicles and creates smooth flow. Example of this effect can be observed in Zackor (1972), and Macdonald (2008) in which capacity was increased by 5% to 10%. The shockwave propagation rate may be reduced when speed harmonization reduces traffic speed and consequently arrival rate at appropriate time and location upstream of a bottleneck. This could shorten the congestion duration and decrease travel time and delay. Examples of this application are reflected in Berton et al. (2002), and Hegyi et al. (2005) in which total travel time was reduced by 12% and 20%, respectively.

Speed harmonization can improve safety by displaying a reduced speed limit based on restricted geometric conditions (e.g., a tunnel), or inclement weather conditions. Also, a speed harmonization system can lower speed of vehicles joining a queue to reduce the likelihood of rear-end collisions. Application of speed harmonization in Germany resulted in 30% reduction in personal injury crashes. Another application of speed harmonization in Netherland reduced collision by 16% (Federal Highway Administration, 2012).

Although successful implementations of this strategy have been reported in the literature, there are some implementations (Lyles et al. 2004, Park and Yadlapati 2002) that resulted in insignificant improvement in traffic operation and safety, and this demands developing more efficient speed harmonization strategies.

In the past studies, the number and location of CMSs were not optimally designed for dynamic speed harmonization. These studies determined signs locations arbitrarily (Hegyi et al. 2005, Ghods et al. 2007, Berton et al. 2002), located the signs next to loop detectors (Allaby et al. 2007), or placed the signs next to on-ramps (Su et al. 2011). None of these approaches necessarily optimize a speed harmonization system to improve traffic operation. Determining optimal location and number of CMSs is important because of the following reasons:

1-It can further improve traffic operation by minimizing travel time or delay. As it will be shown in Chapter 6, there are locations that do not have significant influence on reducing traffic delay. Also there are locations that have some influence but significantly below the optimal level. Possible factors that may lead to the different influences could be distance between a CMS and the bottleneck, as well as different capacities of the roadway segments. As a result, it is important to develop a methodology which considers different factors to find optimal CMS locations.

2- It can avoid installing too many signs and can reduce installation and maintenance costs of the signs.

3- It can improve traffic safety by harmonizing speed. In particular, the number of signs could determine the number of speed changes along the roadway. Eliminating unnecessary speed changes can help to provide smoother speed profile along the roadway and would be more desirable for safety.

Thus, the main objective of this study is to develop and solve a mathematical program, called Work zone Speed Harmonizer (WZSH), to simultaneously find optimal advisory speeds, number of CMSs, and spacing values between the CMSs for work zones. This study is the first study which determines location and number of CMSs when advisory speeds can change over time and space.

Solving the problem is challenging, because it is a large-scale Mixed Integer Nonlinear Program (MINLP). Nonlinearity comes from speed-density relationships and a second order hydrodynamic model (Payne model). Also the program includes integer variables to determine the location of the CMSs. In addition, inclusion of piecewise speed-density models in the program required integer variables. Consequently, this study developed methodologies that tackled challenges coming from integer variables and could be applicable for large scale problems.

The contributions of the study are outlined below. More detailed on the contributions are discussed in Chapter 7:

- The study developed a new methodology to calibrate relaxation time (τ) and anticipation coefficient (ϑ) in the second order model for work zone traffic conditions. The methodology detected the behavior of the second order model in the τ- ϑ space and determined a search direction and its boundaries to avoid stopping at local minima.
- The study proposed the two- ϑ model, which is a modification to the second order model, suggesting two different anticipation coefficients for queue propagation (ϑ_p) and queue shrinkage (ϑ_s). The modification was proposed because the second order model returned slower queue propagation pattern and faster queue dissipation pattern. The two- ϑ model was compared with the second order model using field data and results showed that the two- ϑ model better replicated the queue propagation and dissipation patterns.
- A continuous transformation approach was proposed to eliminate binary variables due to inclusion of piecewise speed-density relationships in the optimization program. Thus instead of solving a mixed integer nonlinear program, which is very challenging to solve for a large scale problem, one can solve the corresponding nonlinear program and significantly reduce computational difficulties.
- The Approximate Decomposition method was proposed to determine optimal location and number of CMSs and dynamic advisory speeds. The Approximated Decomposition method was compared with Greedy Algorithm and Augmented-Cut and Branch method and results showed that the three methods return very close objective function values but the Approximate Decomposition method requires less computational resources.

This study includes 7 chapters. Chapter 1 introduced and presented the objective of the study. Chapter 2 reviewed related studies. Chapter 3 presented the formulation of the problem. One of the formulated constraints is the second order model which has not been calibrated for work zones. A calibration methodology was proposed in Chapter 4 for work zones. This chapter

further proposed a new second order model to improve queue propagation and dissipation patterns. Chapter 5 solved the WZ-Speed Harmonizer when speed-density relationships are piecewise which could be linear or nonlinear. Chapter 6 proposed algorithms to determine optimal locations of signs. Chapter 7 presented conclusions and recommendations.

Chapter 2 Literature review

In this chapter, first articles on speed harmonization are reviewed. Then the models that are used to estimate traffic state in bottlenecks and then methods of capacity estimation for work zones are presented. At the end, conclusions are provided on how the past studies determined number and locations of changeable message signs and what issues should be addressed in the current study.

2.1 Speed harmonization

This section divides speed harmonization studies into work zone and non-work zone studies and then further divides them into field studies and simulation studies. At the end of this section, the effect of speed harmonization on traffic operation will be discussed.

2.1.1 Field studies for speed harmonization in work zones

A variable speed limit control scheme was developed and tested in a 2.5-mile work zone near Wakota Bridge on I-494 by Kwon et al. (2007). Three variables speed limit signs were placed in the work zones as shown in Figure 2-1. Speed and volume data were collected by radar sensors. "The system is designed to lower the speed of the upstream traffic approaching the work zone bottleneck to the same level as the downstream flow." The results showed about 30% reduction in 1-min maximum speed (i.e. maximum speed data aggregated over 1-min intervals), and 7% increase in total throughput.

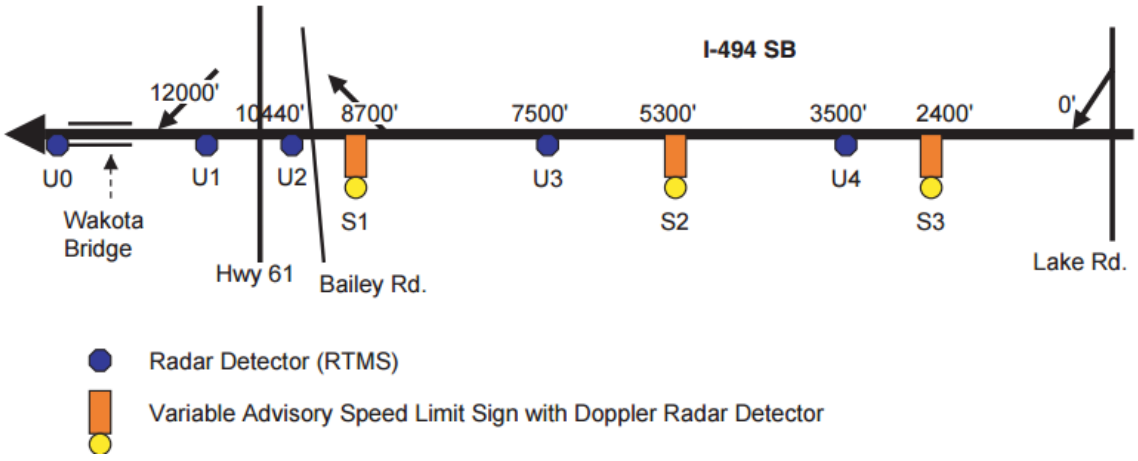


Figure 2-1: Work zone sketch and location of speed limit signs in Kwon et al. (2007)

McMurtry et al. (2008) tested two VSL signs over a six-mile section of a work zone in Utah. Two VSL signs were used and one sign was installed in advance of the work zone and the other sign was placed inside the work zone. The exact spacing between the signs was not reported. Speed data were collected to compare impacts of VSL control with the base condition. In the base condition, a static speed limit of 65 mph was posted in the field, while in the VSL control conditions the speed limit was 55 mph for daytime and 65 mph for nighttime. The results showed that in daytime, speed variation was reduced at the beginning of the work activity area; and in nighttime the average speed was reduced at the same location.

Lyles et al. (2004) evaluated VSL on I-96 in Michigan where "permanent and transient" work activities occurred within an 18-mile work zone. The number of VSL signs was not indicated in the study, but the spacing varied between "a few hundred yards to approximately 1 mile". It was mentioned that due to communication problems between the signs the spacing had to be shorter than the original plan. The displayed speed limit was equal to 85th percentile speed at the next downstream location. Effects of VSL on speed distribution and travel time were assessed, and it was found that VSL control had "minor effects" on 85th percentile speed and speed variance in the work zone and it increased average speed by 1 to 3 mph.

2.1.2 Simulation studies for speed harmonization in work zones

Lin et al. (2008) developed two real-time algorithms to apply variable speed limit control. The first one tries to keep the arrival rate under a specific threshold by reducing traffic speed. This would allow vehicles in the transition area to have enough gaps to merge into the work zone. The second algorithm finds speed limits such that the queue length" in the advance of work zone" does not exceed a threshold. The algorithms inputs are speed and flow data and it is assumed that they can be collected using sensors. The algorithms were applied to 2-to-1 (i.e. One of the two lanes is open in the work zone) , 3-to-1 and 3-to-2 work zones, simulated in CORSIM. The study used five VSL signs, but the spacing between the signs were not reported. The results showed that under "normal traffic condition" the algorithm increased throughput, but the paper did not provide more details on the amount of improvement.

Three control logics to regulate traffic speed were presented by Park and Yadlapati (2002). Both Logic 1 and 2, reduce the difference between average speed of vehicles in the merge area and the average speed of the arriving volume on freeway. In particular, Logic1

computed the average speed of vehicles in the merge area during a given interval and assigned the value to the upstream VSL sign for the next interval. On the other hand, Logic 2 gradually reduced the speed of arriving traffic using two VSL signs. The first VSL sign shows a speed limit, 10 mph higher than the average speed of the merge area, while the second VSL sign displays the average speed of the merge area. Logic 3 used a safety measure, called Minimum Safe Distance Equation (MSDE) to control the traffic speed. A decrease in a value of MSDE is not desirable for traffic safety. Thus Logic 3 reduced (increased) displayed speed limit when MSDE decreased (increased). The study used two VSL signs to control speed in the work zone and located them 2950 ft and 980 ft upstream of the transition taper. All three logics were tested using VISSIM and it turned out that simple application of a static speed limit sign of 45 mph is better than the other logics.

2.1.3 Field studies for speed harmonization in non-work zone bottlenecks

Bertini et al. (2006) evaluated the effect of a variable speed limit system on an 18-km German autobahn. Ten CMSs installed with varying spacing values which were not reported. Also the logic for installation of CMS was not discussed. The signs showed a variables speed limit between 60 km/hr to 120 km/hr with increment of 20. The purpose of the speed harmonization was to keep a continuous flow with a speed between 30 km/hr and 40 km/hr when there is a queue in the field. When there was a "stable" condition speed of 80, 100, or 120 km/hr were displayed. When there was an "unstable" condition, the displayed speed was 60 km/hr. The study claimed that there is a strong correlation between traffic speed and displayed speed limit indicating that variable speed limit had some influence on traffic; this was concluded based on the observation of 1-min traffic speed data that changed based on the displayed variable speed limit; however there was no computation for correlation coefficient or other statistical measures. The study mentioned that "even with the reduction of speed limits, shockwaves propagating a distance of 18 km were observed".

Chang et al. (2011) implemented and evaluated a variable speed limit system on a 4.5 mile segment of MD-100. The system had two VSL signs with a spacing of 1.1 miles. The distance between the downstream CMS and the bottleneck was 1.9 miles. The purpose of speed harmonization was to create a smooth speed profile between a segment with free flow speed and the segment with congestion. The field data during one hour of peak period showed that there is

a 25% reduction in travel time. Also total throughput increased by 230 vehicles (the segment had two lanes) during one hour of a peak period.

M25 Controlled Motorways program was evaluated by the Highway Agency (currently known as Highways England) in 2004. The program implemented variable speed limit on a 4-lane highway between Junctions 10 and 15 in London. The program displays speed limits of 50 mph or 60 mph depending on the traffic volume thresholds and no more details on the thresholds were reported. The dynamic speed signs were placed every 1km. Field data showed the system resulted in 6% reduction in "stop-start" condition, 3% reduction in occurring flow break down, and 10% reduction in injury accident.

Bham et al. (2008) evaluated a variable speed limit system over a 38-mile section of a highway. The highway was divided into 6 segments and one CMS was deployed on each segment. The spacing between the CMSs ranged from 6.13 miles to 8.8 miles. The system displayed a speed limit less than 60 mph if based on 30-sec data average occupancy is more than 7% for a lane, average lane volume is more than 10 vehicles, and average speed is less than 60 mph. In this case, the displayed speed is equal to traffic speed in the congested segment rounded up to the nearest multiple of 5 mph. The study only evaluated the effect of variable speed limit only for 4 segments (out of 6). Standard deviation of speed during each 30-min interval was computed and tabulated for before and after deployment. The authors stated that "speed homogeneity can be inferred from the results of the table. Increase in speed homogeneity may cause reduction in traffic crashes which is a positive outcome of the VSL system." As a result, it was concluded that the variable speed limit could improve safety.

2.1.4 Simulation studies for speed harmonization in non-work zone bottlenecks

Alessandri et al. (1999) developed an optimized control approach to design variable speed limit. A closed loop control approach was designed such that the current traffic state is estimated using Kalman filtering approach. Speed dynamics was predicted using METANET model. The Powell's optimization method was used to determine optimal solution. The method determines a search direction and corresponding step by solving an optimization sub-problem. The method was implemented on a 5-km hypothetical freeway where signs located every 1 km. It was shown that the variable speed limit reduced throughput from 6711671 vehicles to 6672573 for analysis duration of 3000 seconds.

Yang et al. (2013) developed an optimized VSL method and evaluated in a simulation model. A Kalman filtering method was developed to estimate traffic speed using simulation data. An optimization program was solved to determine variable speed limits. Two VSL signs were used because there were two locations caused congestion. Each sign was placed about 800 ft upstream of the congestion location. The results showed that the system resulted in 17.4% travel time reduction and 42.4% reduction in number of stops. The objective function minimized the speed variance.

A VSL control model was develop by Hegyi et al. (2005) for highways. The objective function was to minimize the total time spent in the network. A 12-km stretch of highway was divided into 1-km segments. Speed and density of each segment was updated using the METANET model. Speed limit signs were located in six consecutive segments with a spacing of 1 km. The advisory speed limit upstream of bottleneck was determined to minimize the total time spent in the network. The algorithm was implemented for a benchmark problem resulting in 20% reduction in total travel time.

Allaby et al. (2007) used PARAMICS to evaluate a decision tree based algorithm for speed harmonization on freeways to provide safer deceleration for vehicles joining queue and to increase traffic speed in bottleneck by reducing stop-and-go conditions. Thirteen VSL signs were installed next to loop detectors with a spacing of 500 m to 600 m.The following figure shows the original algorithm. The thresholds were selected based on engineering judgment.

Based on the algorithm if occupancy is less than or equal to 15%, the speed limit will be the normal speed limit. If occupancy is more than 15% and volume is less than 1600 vphpl, the speed limit will be equal to the average traffic speed. To evaluate this algorithm a safety measure called crash potential measure was calculated based on three variables 1) coefficients of variation of speed, 2) the spatial variation of speed, and 3) the covariance of volume. The algorithm was implemented for peak, near-peak and off-peak traffic conditions and resulted in up to 39% reduction in crash potential measure at the expense of up to 23% increase in travel time. It was shown that the excessive travel time can be reduced from 23% to 13% if the occupancy threshold of 20% and volume threshold of 1800 vphpl are used.

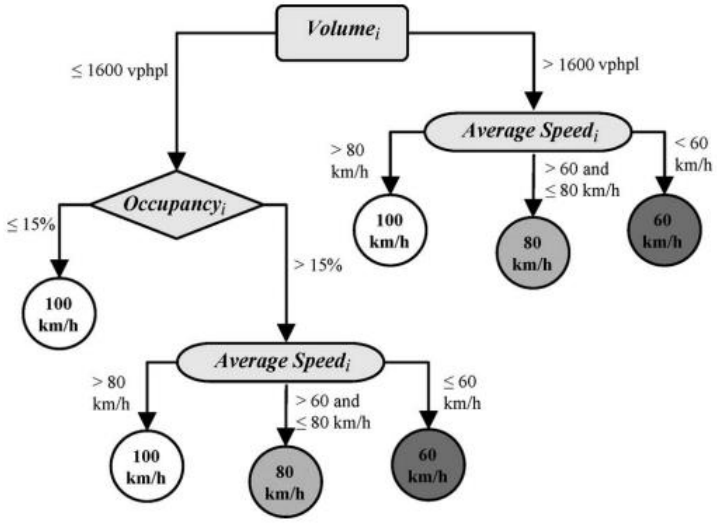


Figure 2-2: Decision tree algorithm for variable speed limit (Allaby et al. 2007)

Ghods et al (2007) proposed an integrated ramp metering and variable speed limit control using fuzzy logic. The if-conditions were based on speed, flow and queue occupancy rates on the main line and the then-statements were the rate of metering and the displayed speed limit. The membership functions were tuned using genetic algorithm and the algorithm was implemented to minimize the total time spent in the network. The METANET model was used to find the travel speeds. The algorithm was tested in a benchmark problem for a 6 km freeway section. Three VSL signs were used with 1km spacing upstream of an on-ramp and resulted in about 6.5% reduction in total travel time.

In another study, Ghods and Rahimi-Kian (2008) proposed a game theory approach to design coordinated ramp metering and variable speed limit. Speed limit signs were considered as players having the same interest. The problem was broken into subproblems, each of which was finding the speed limit for each sign as a player of the game. The results were simulated using METANET model for a 10 km freeway section, with two onramps, and two VSL signs. Both VSL signs were installed upstream of the first ramp with a spacing of 1 km. It was found out that total travel time was reduced by 16%. In addition, the speed limit can be determined fast enough for real time application and better results compared with no control scenario.

Piao and McDonald (2008) evaluated safety impacts of variable speed limit on freeways using micro simulation. The results showed that variable speed limit reduced the percentage of

short headways by an unspecified amount. This study assumed that all vehicles are equipped with in-vehicle speed limit display.

Kamel et al (2008) designed an integrated ramp metering and variable speed limit control. The shockwave propagation and traffic speed were estimated by solving the conservation equation. The conservation equation for on and off ramps was solved using flatness method. The flatness method is an approach to solve differential equations. Using the flatness method, inputs and state variables in the differential equations can be converted to algebraic equations with no partial differential equations. How to solve the equations and the effect of the integrated system were illustrated through an example and the results showed that the integrated system is superior than the "ramp metering-only' system. In particular, the maximum number of vehicles in queue was reduced from 300 to 200.

Lu et al (2010) minimized total time spent in the system using Variable Speed Limit system combined with ramp metering. Assuming a given ramp metering set-up speed limits were determined using Model Predictive Control and METANET model was modified for a general speed-density curve. The algorithm was evaluated in a simulation model for a 2.7 mile segment of I80 EB, in California. Two CMSs were used. One was located about 500-600 meters upstream of the bottleneck. The exact spacing between two signs was not reported. For 5 hours of analysis the VSL system reduced total travel time from 14000 hr to 5300 hr.

Wang (2011) developed a macroscopic model to predict speed dynamics in traffic control problems. The model is a combination of the second order speed-density model and a car following model. When vehicle speed is less than the speed limit, traffic speed is derived from the second order model and when it is above the speed limit drivers slowdown in response to the speed limit and their speed in the next time step is estimated using the car following model based on the difference between the current speed and the speed limit. The developed model was compared with METANET and VISSIM models for a hypothetical 5-lane unidirectional freeway with 5 VSL signs and a spacing of 500 meter. Under different hypothetical speed limits, traffic speed was estimated using the three models. VISSIM estimates were considered as the base and it was concluded that the developed model better represented VISSIM data than the METNET model. No goodness of fit measure was reported for this conclusion.

Su et al (2011) proposed a combined ramp metering and VSL control scheme for weaving sections. The VSL control strategy keeps the occupancy at the critical level for the

discharge section by displaying the critical speed as speed limit. For this purpose, a VSL sign was suggested to be installed at 500 m to 700 m upstream of the bottleneck to provide an appropriate acceleration distance. On the other hand speed limit gradually decreases when vehicles reach the back of queue. The ramp metering control strategy minimizes delay and maximizes travel distance. The Aimsun simulation model for a 6-lane freeway with 7 on-ramps and 7 off-ramps showed that the system improves the system performance when deriver compliance is beyond 30%.

Lu et al (2011) developed a combined ramp metering and VSL control strategy. First the ramp metering rate was determined by minimizing travel time and maximizing mainline flow. Then VSL strategy was designed such that it creates a discharge point before the bottleneck. Within queue speed limit was equal to the critical speed and free flow speed gradually decreased to the critical speed at the back of queue. The paper compared the combined strategy with status quo, ramp metering only, and VSL-only for a 10-km freeway modeled in Aimsun. It was found out that the VSL-only and VSL combined with ramp metering significantly reduced delay, but specific numbers were not discussed in the paper.

Kwon et al (2012) designed a VSL strategy to avoid sudden deceleration. The beginning of the control zone was determined when speed is less than 55 mph for the current interval and deceleration rate is less than or equal to 1500 mile/hr, for three consecutive intervals. These thresholds were determined by observing 20 incidents during a 3-month period and it turned out that half of them happened when downstream speed is less than 55 mph and deceleration rate is less than or equal to 1500 mile/hr. The displayed speed limit was determined based on the distance between the signs and deceleration rate of the traffic. The system status was updated every 30 sec. VISSIM simulation results showed that the system reduces sudden deceleration with relatively small increase in travel time. The authors discussed that the minimum spacing between the CMSs should be selected such that it provides a comfortable acceleration or deceleration rate. No more discussions were provided on how the number of signs and corresponding spacing values were selected in this study.

Hadiuzzaman and Oiu (2012) incorporated capacity drop in cell transmission model to design a VSL strategy for a lane drop. Model Predictive Control was used to determine the speed limits and it was evaluated for an 11-km urban freeway in Canada. Two VSL signs were used. One sign (critical VSL) is placed 200-250 meter upstream of the bottleneck and the second sign

was placed two cells upstream of the bottleneck. The cell lengths were 800 meter, thus the spacing between the second CMS and the bottleneck was 1.6 km. It was stated that the second sign was considered for "keeping the queue length smaller that might grow from the critical VSL. The purpose is to avoid blocking of the upstream on-ramp". The results showed using one sign could improve flow up to 7.1% while using two signs improved flow up to 6.9%.

Zhang et al. (2015) determined optimal location and number of CMSs using Genetic Algorithm. It was assumed that the speed limit for a specific sign is fixed and does not change over time. Cell transmission model was used to model traffic dynamics. A numerical example was solved for a 9-km freeway section for two purposes. The first purpose was to reduce a smoothness factor which was defined as the difference between the densities of two consecutive cells. It was assumed that maximum number of signs is 8. The program determined 6 signs with spacing values ranged from 631 m to 2922 m. The "smoothness measure" was reduced by 87%. The second purpose was to reduce environmental impact, in particular "total human emission exposure". For this purpose, the optimization determined two signs. The most downstream sign was located 4160 ft away from the bottleneck and the spacing between the two sign was 4840 ft. The program could reduce total human emission exposure by 25%.

2.2 Effect of speed harmonization on traffic operation

Two issues are discussed in this section: 1) effect of speed harmonization on capacity 2) effect of speed harmonization on shockwave propagation

1) Effect of speed harmonization on capacity: Field studies demonstrated both significant and insignificant effects of speed harmonization on capacity. An early field study conducted by Zackor (1972) showed that variable speed limit increased capacity by 5-10% in a German Autobahns. Later, Macdonald (2008) reported 7% capacity increase in an England motorway. On the other hand, Papageorgiou et al. (2008) and Smulders (1990) reported insignificant effect of speed harmonization on capacity for European motorways. The mixed effect of speed harmonization on "maximum flow prior to breakdown" (i.e. capacity) was observed by Kianfar et al. (2012) for 8 highway work zones. The study reported 5% to 9% increase for 3 locations, no effect for one location, and 2% to 7% reduction for 4 locations.
2) Effect of speed harmonization on shockwave propagation: Speed harmonization can reduce shockwave propagation by reducing traffic speed and consequently arrival rate. In this approach, which will be used in the current study, traffic speed is determined as the

minimum of two values: a) speed which is read from a speed-density relationship b) advisory speed (Hegyi 2005). This approach was evaluated in benchmark problems resulting 12% to 20% reduction in travel time when there is queue due to a capacity drop in a bottleneck. (Berton et al. 2002, Hegyi et al. 2005).

Based on the discussion above, effects of a speed harmonization system and its design on capacity and fundamental diagram should be further understood and supported using field data Papageorgiou et al. (2008).

2.3 Traffic state estimation

Traffic dynamics models are used to describe spatial and temporal changes of speed, density, and flow rate. Two types of model are introduced in this section: 1) LWR (Lighthill and Whitham 1955, Richards 1956) model, and 2) Second order continuum model. At the end of this section, conceptually the two models are compared. Numerical comparisons using field data will be provided in Chapter 4.

2.3.1 LWR (first order) model

Lighthill and Whitham (1955), and Richards (1956) showed that the solution of conservation law for traffic flow leads to definition of shockwave which is defined as the discontinuity between traffic states. Using this model, one can compute shockwave speed to predict location of shockwave along the roadway and then estimate extend of congestion. Shockwave speed U_{sw} can be computed as:

$$U_{sw} = \frac{\Delta q}{\Delta k} \qquad 2\text{-}1$$

Where Δq is the difference between flow rates at the two sides of the discontinuity. Similarly, Δk is the difference between densities at the two sides of the discontinuity.

In the traffic control problems finite difference form of the conservation law is used to numerically approximate extend of congestion. Assume that the roadway is divided into sections with length of L miles, and time interval duration is Δt sec. It is intended to determine the state of the system for the section i at the beginning of time interval t+1. It is also assumed that the state of the system in the previous time interval is known. The following three equations are used to determine average speed ($V_{i,t+1}$), density per lane ($D_{i,t+1}$) and exit flow rate per lane ($F_{i,t+1}$) for the section i in the beginning of time interval t+1 (Papageorgiu et al. 1989). If λ_i denotes the number of open lanes for section I, then

1) Conservation law

$$D_{i,t+1} * \lambda_i * L = D_{i,t} * L * \lambda_i + \left(F_{i+1,t} * \lambda_{i+1} - F_{i,t} * \lambda_i\right) * \Delta t \qquad 2\text{-}2$$

It should be noted that the section i+1 is at the upstream of section i.

2) Updating exit flow rate

$$F_{i,t+1} = \overline{\alpha} * Q\left(\min\{D_{i,t+1}, D_{cr}\}\right) + (1 - \overline{\alpha}) * Q(D_{i-1,t+1}) \qquad 2\text{-}3$$

$$\overline{\alpha} = \begin{cases} \alpha & \text{if } D_{i-1,t+1} < D_{cr} \\ 1 - \alpha & \text{if } D_{i-1,t+1} \geq D_{cr} \end{cases}$$

Where $Q(D_{i-1,t+1})$ is the flow rate corresponding to $D_{i-1,t+1}$, and D_{cr} is the critical density (i.e., density at the maximum flow rate). The model parameter is α which in theory is equal to 1; in practice values that are, less than but, close to 1 have been obtained based on calibration using field data (Papageorgiu et al. 1989, Kwon et al. 2007). The section i-1 is at the downstream of the section i.

3) The relationship between the traffic parameters

$$F_{i,t+1} = D_{i,t+1} * V_{i,t+1} \qquad 2\text{-}4$$

There are some drawbacks in using the first order model. As discussed by Papageorgiou (1998), the first order models, in general, cannot reflect capacity drop due to lane drop, and ramp metering effect to increase mainline flow rate. Also the first order model assumes that speed change occurs instantaneously and as a result there is no shock structure. To overcome these drawbacks second order model was proposed to more realistically describe asymmetric acceleration and deceleration process. A comparison between first and second order models will be presented in Chapter 4 using work zone field data.

2.3.2 Second order continuum models

The second order models could be in the form of a flow dynamics (Lighthill and Whitham 1955) equation or a speed dynamics equation (Payne, 1971). The speed dynamics form has been frequently used in traffic control problems.

In order to update the state of system, one can use the Equations 2-2 , 2-4 and the following equation which is known as the (extended) Payne model: (Papageorgiu et al. 1989, Breton et al. 2004, Ghods et al. 2007).

$$V_{i,t+1} = V_{i,t} + \frac{\Delta t}{L} V_{i,t}\left(V_{i+1,t} - V_{i,t}\right) + \frac{\Delta t}{\tau}\left(V(D_{i,t}) - V_{i,t}\right) - \frac{\Delta t * \vartheta}{\tau * L}\left(\frac{D_{i-1,t} - D_{i,t}}{D_{i,t} + \kappa}\right) \quad 2\text{-}5$$

Where $V(D_{i,t})$ is the speed at density of $D_{i,t}$ and is computed from speed-density relationship.

The model includes 5 parameters: L, Δt, τ, ϑ, and κ. L and Δt are usually determined before model calibration. The parameters L and Δt should be determined such that it follows the following rule (Papageorgiu, 1998); otherwise the simulation would not be numerically stable.

$$\frac{L}{\Delta t} \geq V_{max} = \text{Maximum traffic speed} \quad 2\text{-}6$$

On the other hand, L and Δt should not be very large as they cause some inaccuracy in estimating the locations of back of queue. Table 2-1 shows that except the early study in 1976, values of $\frac{L}{\Delta t}$ fall between 144 $\frac{\mathbf{km}}{\mathbf{h}}$ and 180 $\frac{\mathbf{km}}{\mathbf{h}}$ (90 mph and 112.5 mph).

The other three parameters calibrated using field data are:

1) <u>The relaxation time (τ)</u> which is the time that a vehicle react to speed changes and expected to be in order of 1-1.8 seconds (Revised Traffic Flow Monograph, Revised Chapter 5). However the calibrated parameters are larger than recommended values. As shown in Table 2-1, the range is from 8 seconds to 72 seconds.

Table 2-1: Calibrated parameters for Second order models

Study	**L (meter)**	**T (Sec)**	**τ (Sec)**	$\frac{\mathbf{L}}{\mathbf{T}}$ $(\frac{\mathbf{km}}{\mathbf{h}})$	ϑ $(\frac{\mathbf{Km^2}}{\mathbf{h}})$
Grewal and Payne (1976)	0.8 (0.5 mile)	4.5	8	640	3.84 $(1.5 \frac{mile^2}{hr})$
Cremer and Papageorgiou (1981)	500-550	10	36	180	21.6
Papageorgiou et al (1990)	110-130	2.5	72	158.4	28
Papageorgiou et al (1990)	≈500	10	36	180	35
Kotsialos et al (2002)	400-800	10	18	144	60

2) <u>The anticipation coefficient (ϑ)</u> which shows how drivers adjust their speed with respect to density changes at the downstream section. Payne (1971) suggested:

$$\vartheta = -\frac{V'}{2} > 0 \quad 2\text{-}7$$

Where V' is the derivative of speed-density relationship with respect to density; however in the traffic control problems, the anticipation coefficient is treated to be constant, and it ranges from 3.84 $\frac{Km^2}{h}$ to 60 $\frac{Km^2}{h}$ as shown in Table 2-1.

3) The fine-tuning parameter (κ) which was not present in the original Payne model (1971). Later it was added by Papageorgiu et al. (1981) to avoid unreasonable speed changes when $D_{i,t}$ in the denominator of the last term is very small.

The fine tuning parameter is not the only extension of the original Payne model. Papageorgiu (1998) suggested to add the following constraint to avoid negative speed values or very low speed values that are not expected.

$$V_{i,t+1} = \max\left\{V_{i,t} + \frac{\Delta t}{L} V_{i,t}\left(V_{i+1,t} - V_{i,t}\right) + \frac{\Delta t}{\tau}\left(V(D_{i,t}) - V_{i,t}\right) - \frac{\Delta t * \vartheta}{\tau * L}\left(\frac{D_{i-1,t} - D_{i,t}}{D_{i,t} + \kappa}\right), V_{min}\right\}$$

In addition, terms were added to take into account the disturbance due to presence of on ramps, off ramps, and lane drops (Papageorgiu et al., 1998).

2.3.3 Discussion: The first order models versus the second order models

The first order models, especially in the form of Cell Transmission Model, have been widely used in different traffic control or modelling problems because of simplicity and low computational cost; however this happened at the expense of some simplifications which could be important for some applications. Papageorgiou (1998) discussed applications of first order and second order models for traffic control problems and he stated that the first order models, in general, cannot reflect capacity drop due to lane drop and ramp metering effect to increase mainline flow rate.

On the other hand, there are some discussions that the second order models may reflect some unrealistic features. In particular, Daganzo (1995) pointed out that the second order models may return negative speeds or flows when vehicles reach the end of queue. Then Papageorgiou (1998) proposed a practical solution which restricts flow by zero, and later Aw-Rascel (2000) and Zhang (1998) proposed new second order models that inherently do not produce negative speeds and flows. The second order models can potentially reflect more realistic traffic features and have been used in traffic control problems, but these models need to be calibrated using field data, and in fact there are limited studies for model calibration.

Both the first order models and the second order models have been used in many traffic control problems. One may select a model based on the purpose of the control problems. For instance, in large scale problems, CTM may be used because of simplicity. However, one may use other traffic models if accuracy in estimating shockwave speed is important.

2.4 Work zone capacity

Researchers did not use a unique definition to estimate work zone capacity. The following definitions have been used in previous literature:

- "The discharge flow when there is a continuous flow of traffic." (Benekohal et al., 2004)
- "The traffic flow rate just before a sharp speed drop followed by a sustained period of low vehicle speed and fluctuating traffic flow rate." (Jiang, 1999)
- "The mean queue discharge flow rate from the bottleneck that was located at the end of the transition area." (Al-Kaisy et al., 2000)
- "95th percentile value of all 5-min within-a-queue" flow rate (Dixon et al., 1996).
- "The average volume of the ten highest volumes immediately before and after queuing conditions" (Maze et al., 2000)

Some of these definitions (Al-Kaisy et al., 2000 and Maze et al., 2000) are based on the mean traffic flow rate whereas there are definitions based on the maximum observed values (Dixon et al., 1996). Some definitions return queue discharge rate (Dixon et al., 1996 and Al-Kaisy et al., 2000) while there are definitions (Jiang, 1999 and Maze et al., 2000) based on flow rates that can be processed before and after flow breakdown.

2.4.1 Factors affecting capacity of work zones

Traffic, geometric, construction and environmental characteristics of a work zone may affect its capacity. Researchers tried to quantify the effects of these parameters. This section recaps the main points of recent studies.

Dixon et al. (1996) investigated 24 short-term work zones located in urban and rural freeways. Data were collected from different types of work zones including 2-to-1 lane, 3-to-1 lane, 3-to-2 lane, and crossover for a freeway with two lanes in each direction. Five-minute traffic data, including number of vehicles, distribution of vehicles across lanes, percentage of heavy vehicles, and average speed of vehicles were recorded. Authors defined capacity for the end transition area as "95th percentile value of all 5-min within-a-queue observations" and

reported it in passenger car per hour per lane (pcphpl) by using PCE factors proposed in HCM 94 (PCE for trucks on level terrain was 1.5). Analyses showed that behavior of drivers in queue conditions were the same during day and night. Variations in capacity for the work activity area were larger than the end-transition area. In rural work zones, mean capacity for the work activity area was smaller than the end-transition area. There was no significant difference between these values for urban work zones. Mean capacity of the transition area for rural work zones (1454 pcphpl) was less than urban work zones (1743pcphpl).

In the research conducted by Jiang (1999), traffic volume, speed and vehicle type were collected from four work zones that experienced the queue conditions. Using traffic counters with road tubes, traffic data were recorded at five-minute intervals for "high-traffic-volume" and at one-hour intervals for "low-traffic-volume". These devices were located before the transition area, in the transition area, and close to the work activity area. Free-flow traffic, merging traffic, and work zone traffic were extracted from the raw data. The work zone capacity was defined as "the traffic flow rate just before a sharp speed drop followed by a sustained period of low vehicle speed and fluctuating traffic flow rate." To detect the capacity, time series diagrams for speed and flow rate were plotted in one graph. Based on the capacity definition, Jiang determined the capacities of these sites as well as the queue discharge rates. The capacities of these sites were not statistically different from each other. The mean queue-discharge rate was statistically lower than the capacity. The results were reported in pcphpl, using PCE factors suggested by HCM 1994. (PCE for trucks on level terrain was 1.5).

In order to analyze a work zone located on an interstate highway, Maze et al. (2000) used 15-minute traffic data including volume, speed, and density. In order to compute the queue discharge flow rate, the queue length was recorded in each minute using delineators that were placed at the side of the road. The queue length was reported to the nearest 0.05 mile. Time series diagram for the speed and flow rate data indicated that there was no significant difference in flow rate just before and after queuing conditions. So "the average of the ten highest volumes immediately before and after queuing conditions" was considered as the maximum capacity of lane closure. Data showed that capacity of a work zone located in a rural highway varied from 1400 pcphpl to 1600 pcphpl.

Al-Kaisy and Zahou (2000) studied two long-term work zones located on either side of a six-lane freeway. One lane was closed due to construction in each work zone, and both work

zones had a lane drop of 3-to-2. Lane closures were located at the median and shoulder lane in the downgrade and upgrade direction, respectively. Al-Kaisy et al. defined work zone capacity as "the mean queue discharge flow rate from the bottleneck that was located at the end of the transition area". Detectors collected space-mean-speed and occupancy within 20-sec intervals. By aggregation of these data, queue discharge flow rate during each 15-min interval was obtained. The capacity was found to vary between 1750 and 2150 vphpl and the mean capacity was equal to 1943 vphpl. Due to the lack of information about the vehicle type and the percentage of heavy vehicles, the capacity values were not been reported in pcphpl. The work zone capacity during peak hours (2117 vphpl) was statistically greater than those during off-peak hours (1955 vphpl). In addition, variation of work zone capacity during peak-hours was less than during the remaining hours. Variation of weekend capacity was small. The mean work zone capacity for weekend was less than that for weekdays. However, it was mentioned that adverse weather conditions existed during the weekend.

Al-Kaisy and Hall (2000) used 5-min traffic volume data recorded by video from two sites to investigate the effect of darkness on work zone capacity. One of these work zones was 800m in length with right-side lane closure, one lane open during construction, and insignificant grade. The other site was 500m long and lane-closures were placed at both the right and the left side of the roadway where the work zone was located. The middle lane was open for traffic during construction. There was also an upgrade and a down grade section. Ideal lane width, sufficient light during night, presence of an off-ramp at the downstream of the work zones and presence of queue at PM peak hours were the conditions in both sites. Data were recorded during the PM peak hours in weekdays. Capacity reduction due to the darkness for the first and second site was equal to 7.5% and 3.25%, respectively. It should be noted that the capacities were converted from vphpl to pcphpl by using PCE factors proposed by HCM 1997 (PCE for trucks on level terrain was 1.5).

In order to investigate the effect of driver familiarity, Al-Kaisy et al. (2001) selected two long-term work zones located in each direction of an eight-lane freeway. Two of these lanes were closed in each direction due to construction activity. It should be noted that these sites were located over a bridge. The grades of this bridge in upgrade and downgrade sections were equal to 3% and the length of each section was equal to 1 km. Volume, vehicle type and queue presence were recorded at 15-min intervals by detectors in spring season. Other information such as

incidence occurrence, weather conditions and work activity was also recorded. Data were classified into three groups: a.m. peak and p.m. peak for weekdays and weekend data. Capacity values were statistically different for these groups. A population factor of one was assigned to a.m. peak hour traffic. Based on this assumption, population factors of 0.93 and 0.84 were suggested for p.m. peak and weekend, respectively.

2.4.2 Capacity estimation

Benekohal et al. (2004) analyzed 11 work zones located in interstate highways with two lanes in each direction, but one lane was closed. Three sites were short-term work zones and eight were long-term work zones. Four types of data called as general, geometric, traffic and construction data were collected. Location of work zone, type of traffic (inbound or outbound), weather conditions, police presence and flagger presence belong to the general data. Geometric data consisted of lane width, total number of lanes in each direction, number of open lanes, presence of ramps close to work zone, length of lane closure, position of closed lane, and length of work activity area. The following work activity information was collected: type of work activity, number of workers present, number and size of construction equipment, proximity of work activity to the travel lane in use, and traffic control devices used were recorded. Traffic data included the headway and speed of each vehicle in the work zone, volume of traffic, and queue length. Data related to general conditions, geometry, work activity and how they varied during the data collection period were recorded by field observation. In most of the sites, in order to collect the traffic data, two markers apart from each other with an approximate distance of around 250 ft were used. A camera recorded the traffic stream between these markers. In three sites due to heavy traffic stream, markers could not be installed, so two observers for speed data and one observer for queue data (i.e. presence of queue and length of the queue) were used. Data collection period varied from 2 to 4 hours. Vehicle type, time at each marker, and whether a vehicle was in platoon or not was obtained from time coded videotapes. The accuracy of the reading of travel time was within 1/30 second. In order to measure headway of a particular vehicle, the time at which the front bumper of the vehicle passed the more visible marker was recorded. The corresponding time difference between two successive vehicles gives the headway of the following vehicle. The distance between two markers was measured and the actual distance between the two markers that vehicles traveled was computed considering the angle between camcorder and the markers.

Benekohal et al. (2004) defined capacity as "the discharge flow when there is a continuous flow of traffic." In order to have continuous flow of traffic, platooning vehicles which had a spacing less than or equal to 250 ft or headway less than 4 seconds were considered. Queuing condition was observed in three sites. For these sites, the top 15-min intervals were selected to measure the headways whereas for the others, the top 5-min intervals were used. For each site, service capacity was computed as the reciprocal of the average headway. It should be noted that the average speed of platooning vehicles in these periods was very close to the average speed of all vehicles including non-platooning and platooning.

A speed-flow relation was developed based on the field data. The free flow regime was assumed to be horizontal line. HCM 2000 Equation was modified for the transition regime and exponential regression model was fit to the congestion data. Knowing the operating speed of traffic, one can use the speed-flow relation to estimate the capacity under the prevailing conditions. A model was suggested to estimate operating speed when there is no field data. The model finds the operation speed based on the speed reducing effect of non-ideal conditions such as lane width, lateral clearance and work intensity. The suggested capacities varied between 1600 pcphpl and 2200 pcphpl when free flow speed ranged from 40 mph to 65 mph, respectively.

In another study, Benekohal et al. (2010) used more comprehensive field data than the previous study (Benekohal et al. 2004) to develop speed-flow relationships. This study used data for free flow, transition and congested regimes; while the previous study used field data only for congested regime. Also, this study computes capacity based on field data, while the previous study relied on the intersection of congested regime and transition regimes to estimate capacity.

The study proposed the following three sets of speed-flow relationships for work zones: 1) speed limit =55 mph 2) speed limit=45 mph, and 3) speed limit= 45 mph with flagger. The ideal capacities are 1800 pcphpl for speed limit of 55 mph, 1600 pcphpl for speed limit of 45 mph, and 1400 pcphpl for speed limit of 45 mph and flagger. These capacity values may be reduced if there are factors reducing free flow speed. These factors could be non-ideal lane or shoulder widths, presence of work activity or speed control techniques. The study developed an equation to estimate free flow speed as a function of these factors.

Ramezani et al. (2010) developed a methodology to measure capacity using field data. The methodology can be applied when demand is less than, but close to capacity, or when

demand is higher than capacity. The methodology finds the flow rate corresponding to average headway of in-platoon vehicles as the potential capacity. The potential capacity may be observed when all vehicles are in a large platoon. However mostly there are inter-platoon gap between vehicles reducing the observed flow rate at capacity level. Thus the potential capacity is multiplied by a platooning factor, which is less than 1, to take into account the presence of inter-platoon gaps. The study suggested platooning factor of 1 for queuing conditions, 0.95 for long term work zones and non-queuing conditions, and 0.85 for short term work zones and non-queuing conditions.

Sarasua et al. (2004) conducted a study on 22 short-term work zones in South Carolina's interstate highways. All sites had 12-foot lane width and similar taper length. In this study, two video cameras were used to record the traffic volume and vehicle type. These cameras covered "taper and lane closure transition". Speed was measured by radar gun. When the vehicle speeds were greater than 35 mph, the speed data were aggregated over 5-min intervals. When vehicle speeds were below 35 mph, the speed data were aggregated over 1-min intervals. At the same time interval, queue length was recorded in feet. Based on 1-hr data, a linear speed-density relationship was developed with R-square of 0.93. Therefore, the following Greenshields-type model was suggested to represent the variation of flow with respect to speed.

$$q = -2.05\,s^2 + 109.7\,s$$

Where, q is the flow rate in pcphpl and s is the traffic speed in mph. From this model, a value of 1,467 pcphpl was determined as the capacity of 2-to-1 work zones.

Al-Kaisy et al. (2003) analyzed long-term work zones located in six freeways. Based on the collected field data the following model was proposed:

$$C=C_b * f_{HV} * f_d * f_w * f_s * f_r * f_l * f_i$$

where C_b is the base capacity (2000 pcphpl) and C is the work zone capacity (vphpl). Other adjustment factors and their recommended values are as follows:

F_{HV}= Adjustment factor for heavy vehicles which is computed from the relevant formula proposed by HCM 2000. PCE for trucks and buses was suggested as 2.4 for level terrain and 3.0 for 3% upgrade with 1-km length. Linear interpolation was suggested for 1-km long grades which fall between level terrain and +3%. For other grades with similar lengths, an adjustment proportional to the values proposed by HCM 2000 was suggested.

f_d= Adjustment factor due to driver population (1 for weekday peak-hours,0.93 for weekday off-peak hours and 0.84 for weekends)

f_w= Adjustment factor for work activity (1.00 if there is no work activity at site; otherwise, 0.93 is used)

f_s= Adjustment factor for the side of lane closure (1.00 if there is right-side closure, 0.94 if there is left-side closure)

f_r=Adjustment factor for rain (1.00 if there is no rain, 0.95 for light to moderate rain, 0.90 for heavy rain)

f_l=Adjustment factor the effect of the Lighting conditions (1 for daytime and 0.96 for night with good Lighting conditions)

f_i=Interactive effects (1.03 for left-side lane closure during weekday off-peak hours, 1.08 for weekends when work activity is present, 1.02 for left-side closures during weekends, 1.05 for rain during weekends, 1.00 for all other conditions)

To investigate the factors affecting work zone capacity, Venugopal and Tarko (2001) collected data from a long-term work zone with heavy work activity. This work zone was located on a four-lane rural freeway. There was a left-side lane closure in one direction and a right-side lane closure in the other direction of the freeway. The investigated factors were rain, wind, heavy vehicles, the location of lane closure (shoulder or median) police presence and "Indiana Late Merge System (ILMS)". ILMS is a new control device "to improve the use of traffic lanes on work zone approaches". Speed, volume and vehicle type were aggregated over 20-min intervals. To detect capacity conditions, time-series diagram for speed and flow rate was plotted. A sharp drop in speed with a considerable period of low speed indicated capacity conditions. Capacity values varied in a wide range. The mean capacity value was equal to 1320 veh/hr. Analysis of Covariance (ANCOVA) indicated that ILMS, weather conditions (rainy or sunny), heavy vehicles and presence of police were the only effective variables on the work zone capacity at 5% significance level. The highest wind speed recorded during data collection period was equal to 24 km/hr which was not very strong. Based on these results, a multiplicative model is presented as below:

$C=C_o * f_M * f_R * f_H * f_p$

Where:

C= Work zone capacity under existing conditions,

C_o = Work zone capacity under ideal conditions (no ILMS, absence of police, no truck and no rain),

f_M = Adjustment factor for presence of ILMS,

f_R = Adjustment factor for rainy weather,

f_P = Adjustment factor for police presence,

f_H = Adjustment factor for heavy vehicles

Following the calibration of the model, a value of 1440 veh/hr was obtained for C_o and a value of 1.4 was reported for EH (PCE value for heavy vehicles) on level terrain. The values of f_M, f_R and f_P were equal to 0.94, 0.91 and 0.86, respectively.

Kim et al. (2001) developed a new model to estimate capacity for short-term work zones. To reach this point, 12 sites were selected. All sites were located on eight-lane freeways. The following factors were enumerated as the parameters which may affect the work zone capacity: number of closed lanes and open lanes, location of lane closure (shoulder or median), percentage of heavy vehicles, familiarity of drivers with the work zone, presence of on-ramp near the work zone, lateral clearance, length and grade of the work zone, severity of work activity (low, medium, and heavy), the length of time that work zone exists (short-term or long-term), weather conditions, and work time (night or day time).

Traffic data were collected after p.m. and a.m. peak hours, so it was assumed that drivers were not familiar with the sites. To record volume data, a video camera was installed after the end-transition area. A laser speed gun was used to collect speed data at 1-min intervals. Queues were observed only at upstream of work zones. Vehicles were classified into passenger cars and heavy trucks. In addition to the traffic data, characteristics of work zones were also recorded. These characteristics zone and the corresponding capacity values for each work are shown in the next section. Intensity of work activity for each site was indicated with regard to the type of construction, number of workers and size of equipment. Based on the results of the correlation matrix between the factors, which may affect the capacity of work zones, the following model was proposed to estimate work zone capacity:

$$\mathrm{CAPACITY} = 1857 - 168.1\mathrm{NUMCL} - 37.0\mathrm{LOCCL} - 9.0\mathrm{HV} + 92.7\mathrm{LD} - 34.3\mathrm{WL} - 106.1\mathrm{WI_H} - 2.3\mathrm{WG} * \mathrm{HV} \quad (\text{Adjusted } R^2 = 0.98)$$

Where:

CAPACITY = Work zone capacity in vphpl,

NUMCL = "Number of lane closures,"

LOCCL = "Dummy variable for the location of lane closure (in the case of right-side lane-closure, LOCCL = 1; otherwise, LOCCL =0),"

HV = "Proportion of heavy vehicles,"

LD = "Lateral distance to the open travel lane,"

WL ="Work zone length,"

WI_H =" Dummy variable for heavy work activity (for heavy work activity, WI_H=1; otherwise, WI_H=0),"

$WG * HV$ = Work zone grade*proportion of heavy vehicles.

The results indicated that CAPACITY had a high correlation with NUMCL, WI_H, and $WG * HV$.

Karim and Adeli (2003) proposed a model to estimate work zone capacity by using radial basis function neural network (RBFNN). The model was established based on the following factors: number of lanes, number of open lanes, work zone configuration, work zone length, lane width, percentage of trucks, grade, work zone speed, work intensity, darkness and presence of ramp 1500 ft before the transition area or 500 ft after the work zone. In this model, work intensity was determined with regard to the size of equipment, the number of equipment and workers, the amount of dust and noise generated and the distance of the work area to the traveled lane. To obtain data for the training of the model, authors used the capacity values proposed by ODOT for work zones with various characteristics, but before applying them, those data were modified with regard to the HCM 2000 guidelines and engineering judgment of the authors. After that, the model was trained by using 40 examples of work zone capacity. To evaluate the accuracy of the model, nine samples were taken from the North Carolina studies, 12 samples from Indiana studies, and six samples from Maryland studies. However, there was no information for the value of some parameters that affect work zone capacity. Hence, Karim and Adeli assumed a value for the missing information. Comparison of the estimated values with the observed values indicated that the model had a large error in some cases. In all cases with large error, percentage of trucks was high (19-32%). Two possible sources were mentioned for these errors. Firstly, the model can only simulate the effect of trucks on grades but authors mentioned

that trucks had another effect on the capacity of work zones, too. This effect is mean speed reduction due to the presence of trucks which can happen even in level terrain. There was not any information in the training data for this effect. Secondly, there was not any information for grades higher than 25%. A root mean square of 165 (veh/hr) was reported for the model.

Adeli and Jiang (2003) proposed a neuro-fuzzy logic model to estimate work zone capacity. It was assumed that the following parameters affect work zone capacity: percentage of trucks, grade, number of lanes, number of closed lanes, lane width, work zone configuration, work intensity, length of work zone, work zone speed, interchange effects, work zone location (urban or rural), work zone duration (short term or long term), work time (day or night), work day (weekday or weekend), weather conditions, pavement conditions (dry, wet, or icy), and driver population.

They used 168 datasets for testing, checking, and training the model. These datasets were collected from North Carolina, Texas, California, Indiana, Maryland, Ohio, and from Toronto, Canada. There was no information available for some of the variables. Some values were assumed for these variables. To evaluate the proposed model, estimated values versus observed values were plotted on one chart. The chart did not show any outliers. An outlier was defined as "any point with an error value 50% larger than the mean error for all the data points." A lower root mean square was reported for the neuro-fuzzy model compared to the models by Krammes et al. and Kim et al (2001).

2.5 Conclusions

This section presents conclusions for determining number and location of CMSs for speed harmonization as well as the traffic state models

2.5.1 Determining number and location of CMSs for speed harmonization

Review of past studies showed that the studies can be divided into three categories:

Category 1) Studies (Hegyi et al. 2005, Ghods et al. (2007), Berton et al. 2002) that provided no discussions on how to select the number and locations of the CMSs and arbitrarily assumed that CMSs are located with a uniform spacing.

Category 2) Studies that provided some recommendations on choosing locations of CMSs, but the discussions may not necessarily improve or optimize traffic operation. The recommendations are:

- Kwon et al. (2012) recommended that spacing between two consecutive speed limit signs (d_i) should be at least equal to

$$d_i = \frac{SL_i^2 - SL_{i+1}^2}{2a}$$

Where SL_i and SL_{i+1} are advisory speed limits displayed by two consecutive signs and a is the average deceleration rate. No more discussion was provided on the number and location of the CMSs used in that study.

- Allaby et al. (2007) assumed that speed limit signs are installed next to loop detectors. This resulted in 13 signs with spacing of 500-600 meter for an 8-km section of a highway
- Su et al. (2011) assumed that VSL signs will be used where on-ramps merge the mainline. Thus 7 signs were used for a 10-km stretch of a highway.

Category 3) This category includes only one study conducted by Zhang et al. (2015) who developed an optimization program to determine the optimal location and number of CMSs. However, they assumed that variable speed limit does not change over time while there are many studies ((Hegyi et al. 2005, Ghods et al. (2007), Berton et al. 2002, Kwon et al. (2012)), that considered advisory speed as a decision variable that can change over time and space for queuing and non-queuing conditions. This assumption will over simplify the problem, and reduce the capability of speed harmonization. .

The categories above indicate that there is a need to develop an optimization program, like WZSH, to simultaneously determine number and location of CMSs and dynamic advisory speeds.

2.5.2 Traffic state models

Using a proper model to estimate traffic state is important to develop an effective and reliable speed harmonization system. Review of past studies showed that there are two issues that are needed to be addressed in this study 1) there is no study that compared LWR model versus second order model for work zones, and 2) there is no study which calibrated the second order model for work zones. These issues will be addressed in this study and the model that better represent field data will be incorporated in the optimization program.

Chapter 3 Formulation

The formulation is explained for a 2-to-1 work zone (i.e. out of the two lanes only one lane is open to traffic) as shown in Figure 3-1. Traffic moves from right to left, and the bottleneck location is assumed to be at Work Space (WS); thus speed harmonization is implemented within the WS as well as upstream of the WS. The roadway is divided into sections of fixed length L, and, for the purpose of formulation, it is assumed that each section is equipped with a Changeable Message Sign (CMS) at its upstream boundary. The CMSs display advisory speeds with the following characteristics:

- Advisory speeds for a given CMS may change at the end of each cycle, where a cycle is a user-defined time interval during which displayed advisory speed does not change. The magnitude of the advisory speed for each interval is determined by solving the optimization program.
- There are two types of CMSs: 1) critical and 2) non-critical. If a CMS is critical, its displayed advisory speed can be different from the advisory speed for the immediate upstream CMS. Consequently, if a CMS is non-critical, its advisory speed will be always the same as that for the immediate upstream CMS. The number and locations of the critical CMSs are determined by solving the optimization program.

3.1 Assumptions

The formulation is presented for typical 2-to-1 highway work zones. The following assumptions are made:

Assumption 1: Traffic state in the work zone at the beginning of the analysis period is known.

Assumption 2: "Base" speed-flow curves for all the sections are available. "Base" refers to the condition with no speed harmonization.

Assumption 3: The roadway and time are discretized, and L and Δt denote the section length and the time interval duration, respectively. Figure 3-1, shows discretized roadway sections numbered from the downstream to the upstream. The time interval, Δt, is assumed to be equal to the cycle length (Berton et al. ,2002). Cycle length is defined as the duration that displayed advisory speed is fixed and does not change. At the end of each cycle, advisory speed may or may not change. The program includes constraints to control the amount of speed change from one cycle to another.

Assumption 4: Traffic stream speed fully complies with displayed advisory speeds.

Assumption 5: Average speed of vehicles for a section with steady state condition is computed as the minimum of two values: 1) advisory speed, and 2) the speed computed from speed-density relationship, Berton et al. (2002).

The rest of this section illustrates definitions, objective function, and constraints.

Figure 3-1: Discretized Network for the Benchmark Problem

3.2 Definitions

3.2.1 Decision variables

x_i : A binary variable which is defined for Section i to determine the locations of the critical signs. It is 1 when a critical CMS is installed in Section i; otherwise it is zero.

$AS_{i,t}$: Advisory speed that is displayed by a CMS at the upstream boundary of the Section i during the time Interval t.

3.2.2 State variables

$V_{i,t}$: Average travel speed (mph) for the Section i at the beginning of Interval t.

$D_{i,t}$: Density (veh/mi/ln) for the Section i at the beginning of Interval t.

$F_{i,t}$: Flow rate (vphpl) exiting from the Section i at the beginning of Interval t.

$V(D_{i,t})$: Speed corresponding to density $D_{i,t}$ computed from speed-density relationship for the Section i at the beginning of time Interval t. Average travel speed, $V_{i,t}$, may be different from $V(D_{i,t})$ because the "second order" speed dynamics model is used to keep track of traffic evolution.

$U_{i,t}$: Steady state speed for the Section i at the beginning of time Interval t. Based on Assumption 5, it is the minimum of two values: 1) $AS_{i,t}$, and 2) $V(D_{i,t})$.

TTT: Total travel time (veh-hr) during the analysis period.

3.2.3 Parameters

N_i: Total number of sections.

N_t: Total number of time intervals in the study period.

Δt: Time interval duration for updating the state of traffic.

L: Section length.

λ_i: Number of open lanes for Section i.

τ : Relaxation parameter in the "second order" hydrodynamics model.

ϑ: Anticipation parameter in the "second order" hydrodynamics model.

ϕ : Coefficient for the effect of lane drop in the "second order" hydrodynamics model.

D_{cri}: Critical density of the section i (pc/mi/ln).

Δ_t: Maximum change in advisory speeds displayed by a given CMS for two consecutive time intervals. This is considered as maximum temporal changes.

Δ_s: Maximum change in advisory speeds displayed by two consecutive CMSs during a given time interval. This is considered as maximum spatial changes.

$MaxAS_i$: Maximum advisory speed that may be displayed in the Section i.

$MinAS_i$: Minimum advisory speed that may be displayed in the Section i.

m: minimum number of sections between two critical CMSs.

3.2.4 Sets

I: Set of indices for all sections. Each member of this set is denoted by i.

I^*: Set of indices for sections with critical CMSs. This is a subset of set I.

T: Set of indices for all time intervals. Each member of this set is denoted by t.

3.3 Objective function formulation

The objective function is to minimize the total travel time plus a penalty function for the number of signs.

$$\text{Min}((\sum_i \sum_t D_{i,t} * \lambda_i * L * \Delta t) + P(n)) \quad 3\text{-}1$$

The first term is the total travel time and is the same as the objective function used in (Berton et al. 2002, Carslon et al. 2011). The second term is the penalty for the number of critical CMSs; where, n is the number of critical signs and P(n) is the penalty function for number of signs. The penalty function is a user-defined function and depends on the strategies to install the signs. This study uses the following linear penalty function

$$P(n) = p * n$$

Where, p is the marginal "cost" for installation of one sign. The marginal "cost", p, is defined as the average free flow travel time for the entire control zone. (Control zone is defined from the

very upstream section till the end of the work zone). This means that an extra sign is placed if the resulting reduction in travel time is at least equal to the average free flow travel time. The average free flow travel time was selected because it is independent of volume and available capacity. Alternatively one could use average (or a portion of) travel time in base condition but is a function of volume or capacity in the base condition and can change from one day to another day.

3.4 Constraints formulation

There are three sets of constraints:

1) Traffic evolution constraints
2) Safety constraints
3) Other constraints

Each set is formulated as below.

3.4.1 Traffic evolution constraints

The purpose of these constraints is to find average travel speed, exit flow, and average density for a given section at the beginning of the time interval t+1, assuming that all these variables are known at the beginning of interval t.

Constraint 1) Traffic density is estimated using the conservation law:

$$D_{i,t+1} * L = D_{i,t} * L + \left(F_{i+1,t} * \lambda_{i+1} - F_{i,t} * \lambda_i\right) * \Delta t \qquad \forall i \in I, \forall t \in T \qquad 3\text{-}2$$

Constraint 2) Speed dynamics is determined using the "second order" hydrodynamics model proposed by Payne. This model may return negative speed or flow. To overcome this issue, Papageorgiou (1998) suggested that negative flow rates should be adjusted by zero. An implication of this adjustment is that minimum speed should be zero to avoid negative speeds and flow rates. As a result the following model is used:

$$V_{i,t+1} = \text{Max}\left\{0, V_{i,t} + \frac{\Delta t}{L} V_{i,t}\left(V_{i+1,t} - V_{i,t}\right) + \frac{\Delta t}{\tau}\left(U_{i,t} - V_{i,t}\right) - \frac{\Delta t * \vartheta}{\tau * L}\left(\frac{D_{i-1,t} - D_{i,t}}{D_{i,t} + \kappa}\right)\right\} \quad \forall i \in I, \forall t \in T \qquad 3\text{-}3$$

Also, Papageorgiou et al. (1990) suggested the additional term $\frac{-\phi\, T(\lambda_i - \lambda_{i-1}) * D_{i,t-1} * V_{i,t-1}^2}{L * D_{cri} * \lambda_i}$ be included to take into account lane changing disturbance that may occur due to lane drop. It is assumed that this disturbance only occurs within half a mile upstream of the lane drop.

Constraint 3) Constraints 3 finds the steady state speed, $U_{i,t}$, as the minimum of the advisory speed and the speed computed from the speed-density relationship.

$$U_{i,t+1} = \min\{V(D_{i,t+1}), AS_{i,t+1}\} \qquad \forall i \in I, \forall t \in T \qquad 3\text{-}4$$

$V(D_{i,t+1})$ that appears in Constraints 3 could be a piecewise function. How to formulate piecewise speed-density relationships is discussed in (12).

Constraint 4) Exit flow rate is defined as:

$$F_{i,t+1} = D_{i,t+1} * V_{i,t+1} \qquad \forall i \in I, \forall t \in T \qquad 3\text{-}5$$

3.4.2 Safety constraints

Constraint 5) According to definition of a critical CMS, when advisory speeds of two consecutive sections are different ($AS_{i+1,t} \neq AS_{i,t}$) a critical CMS has to be placed for section i. This is imposed by Constraint 5 when x_i is 1. In addition, the constraint controls the spatial changes in advisory speeds. When the sign is critical, the difference between advisory speeds that are displayed by two consecutive CMSs during a given time interval is restricted by Δ_s. This is to avoid a drastic speed reduction between two sections and its potential safety consequences. There is no restriction in increasing the advisory speed to let traffic recover as soon as possible, thus the bound for the speed limit increase is $MaxAS_i$ which is the maximum advisory speed.

$$-MaxAS_i x_i \leq (AS_{i+1,t} - AS_{i,t}) \leq \Delta_s x_i \qquad \forall i \in I, \forall t \in T \qquad 3\text{-}6$$

Δ_s is chosen to be 5 mph in this study.

Constraint 6) This constraint restricts advisory speed fluctuations by Δ_t for a given CMS. Both Constraints 5 and 6 control advisory speed changes, but in different ways. Constraint 5 controls spatial speed reduction for a given time interval, while the Constraint 6 is to control temporal changes for a given sign.

$$|AS_{i,t+1} - AS_{i,t}| \leq \Delta_t \qquad \forall i \in I^*, \forall t \in T \qquad 3\text{-}7$$

Δ_t is chosen to be 5 mph for this study.

Constraint 7) The advisory speeds shall not exceed the static speed limits established based on work zone and freeway design standards. Thus the static speed limit is considered to be the maximum advisory speed ($MaxAS_i$) for each section. The $MaxAS_i$ may change over sections since mostly work zone speed limit is lower than speed limits for basic freeway sections.

$$AS_{i,t} \leq MaxAS_i \qquad \forall i \in I^*, \forall t \in T \qquad 3\text{-}8$$

Constraint 8) This constraint establishes the minimum advisory speed to be $MinAS_i$ for each section. One may use this constraint to avoid very low travel speeds that potentially may be imposed by the advisory speeds. In this study, $MinAS_i$ is set to be 20 mph for all the sections.

$$MinAS_i \leq AS_{i,t} \qquad \forall i \in I^*, \forall t \in T \qquad 3\text{-}9$$

3.4.3 Other constraints

Constraint 9) The following constraint is added to declare the number of the critical signs.

$$\sum_{i \in I} x_i = n \qquad 3\text{-}10$$

Constraint 10) This constraint establishes the minimum spacing between the signs to be "mL" assuming the sections length is L.

$$\sum_{i=j}^{m+j-1} x_i \leq 1 \qquad j \in \{1, \ldots, (N_i - m + 1)\} \qquad 3\text{-}11$$

The formulated problem is a Mixed Integer Nonlinear Program. The integrality comes from the binary variable x_i which determines the CMSs locations. The nonlinearity mainly comes from Constraints 2, 3, and 4.

Chapter 4 Calibration of second order hydrodynamic model for work zones

There are few studies to calibrate the second order hydrodynamic models and they are mostly based on freeway field data and there is no study based on work zone data. Specific study for work zones is needed because traffic operation in work zones is different from that in freeways. In particular, capacity and speed are, in general, lower and drivers are more cautious. Thus this section calibrates the second order model using work zone field data.

4.1 Field data

Data were collected from a 2-to-1 work zone (i.e. one of the two lanes is open in the work zone) on an interstate highway 39 NB. The left lane was closed due to construction activity on a bridge. Speed limit in the work zone was 45 mph and a flagger was present at the beginning of the bridge showing "Slow down" paddle.

Traffic stream was videotaped and its setup is shown in Figure 4-1a. The camera was placed around 200ft-250 ft away from the flagger location where the location of the front of queue was anticipated to be. The camera captured the time that each vehicle passed two markers and the following data were obtained for each vehicle: 1) time headway 2) speed and 3) vehicle type (passenger car or heavy vehicle). Location of back of queue was recorded at the end of every minute by an observer.

The headway accuracy is $1/30^{th}$ of a second; the speed accuracy is 1 mph, and the queue length accuracy is 120 ft.

The reduced data were aggregated over 1 min intervals and the resulting flow rate and average speed are shown in Figure 4-2. The queue length at the end of each minute is also displayed in Figure 4-3. Trucks were converted to passenger cars assuming Passenger Car Equivalency factor is 1.5. This is the PCE value that the HCM 2010 suggests for basic freeway sections. The truck percentages for one-minute intervals range from 0 to 62% with average of 28%.

Forty eight minutes of data, shown in Figures 2 and 3 are used to calibrate the second order model. Data includes pre-break down, break down and post break down conditions. In the pre-break down condition (minutes 1 to 13), speed varies from 26 mph to 38 mph which overall is less than the posted speed limit of 45 mph. The reason is the presence of flagger showing "Slow Down" paddle; as a result vehicles reduced speed in response to flagger. During minute

14, a loading truck parked on the right shoulder, next to the flagger location and some work activity occurred close to travel lane until minute 37. During this time speed varied between 9 mph and 21 mph and queue propagated roughly 45,000? ft upstream of the flagger location. During minute 37 the loading truck left the location and departure speed increased and became more than 25 mph. As a result queue shrank faster and ultimately at minute 43 the queue completely dissipated. In the post break down condition (minutes 44 to 48) speed is roughly between 28 mph and 48 mph.

Number of vehicles arriving back of the queue was approximated for each minute using the data collected from field and the following equation

$$A_{t+1} = D_{t+1} + Q_{t+1} - Q_t \qquad \forall t \in T \qquad 4\text{-}1$$

Where,

A_{t+1}: Number of vehicle arriving back of the queue during the minute t+1

D_{t+1}: Number of vehicles departing queue during the minute t+1

Q_{t+1}: Number of vehicles in the queue at the end of the minute t+1, which is computed as

$$Q_{t+1} = QL_{t+1} * K_{t+1} \qquad \forall t \in T \qquad 4\text{-}2$$

Where,

QL_{t+1}: Queue length at the end of minute t+1 recorded by the observer

K_{t+1}: Density for the minute t+1 computed as the flow rate divided by the average speed for minute t+1

The computed arrival rates are shown in Figure 4-4.

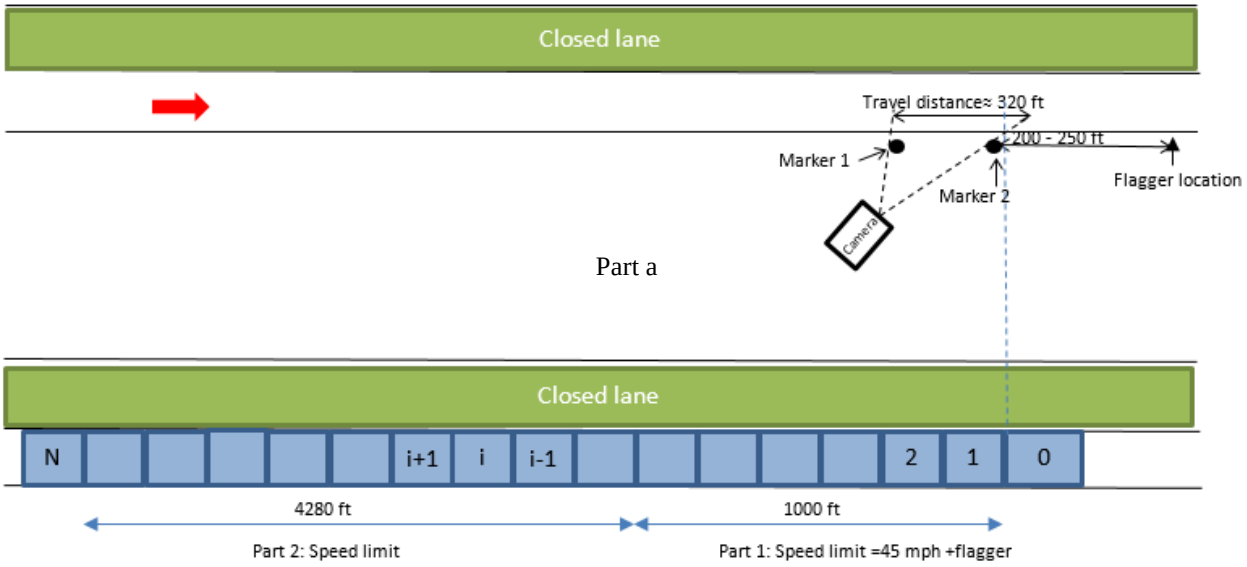


Figure 4-1: a) One lane section of the work zone b) Discretized roadway

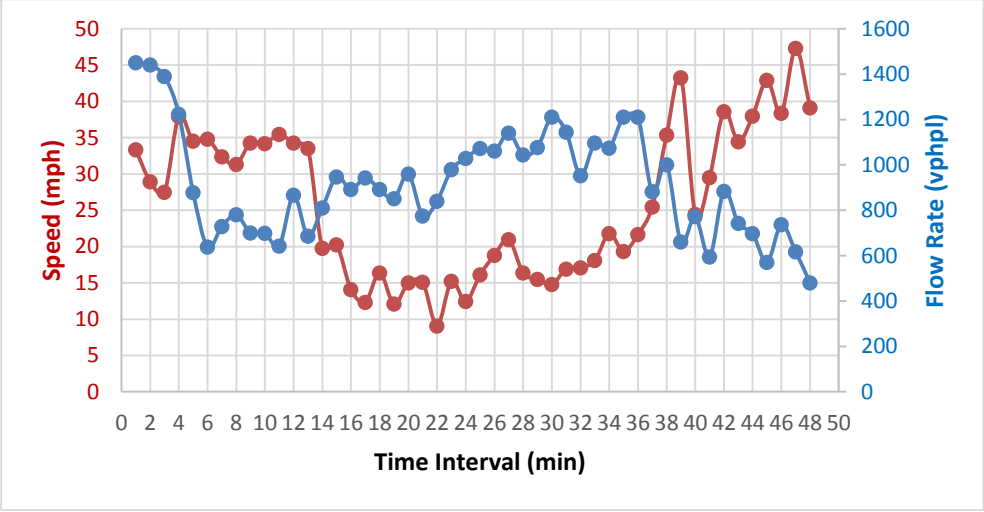


Figure 4-2: One-minute speed and flow rate data

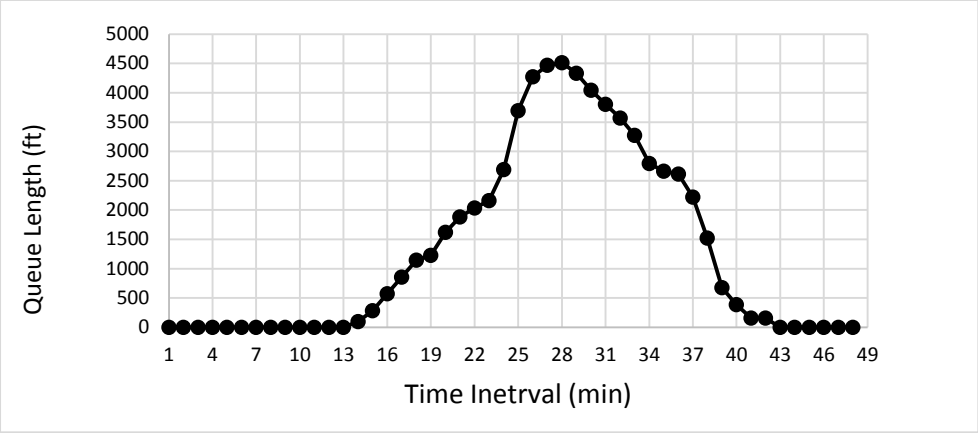


Figure 4-3: Queue length at the end of each minute

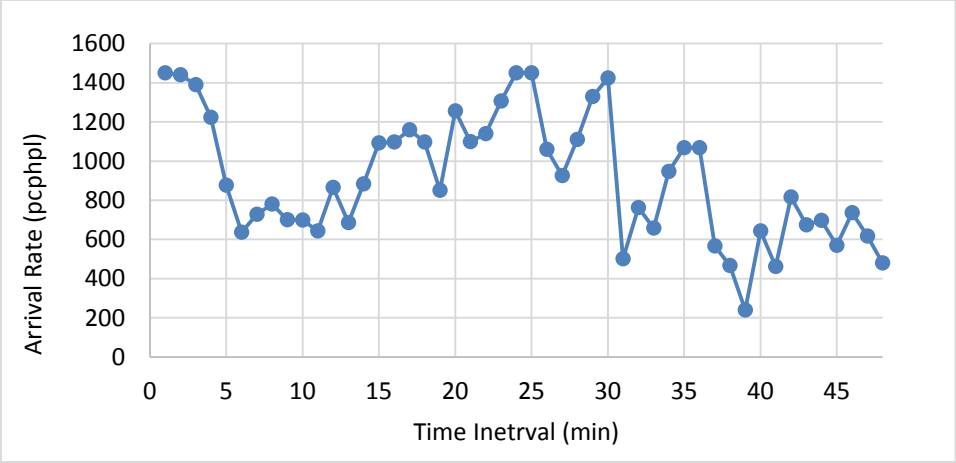


Figure 4-4: Arrival rate approximated based on field data

4.2 Network set up

The roadway is discretized as shown in Figure 4-1b. The sections 0 and N are allocated to impose the boundary conditions and the length of the network between these two sections is 1 mile and includes two parts. Part 1 is 1000 ft long and represents the sections that are influenced by a flagger. In particular, drivers in this part are able to see the flagger and reduce speed in response to flagger. As a result, the fundamental relationship (in form of speed-flow curve), developed by Benekohal et al. (2010), for a work zone with speed limit of 45 mph and a flagger is used. This relationship is shown in Figure 4-5 . This relationship is developed based on the data collected from this work zone and other work zones with similar conditions. Details on how to develop the relationship are explained by Benekohal et al. (2010). Part 2 is 4280 ft long and it is assumed that driving behavior is not influenced by the flagger. For this part, the curve with

speed limit of 45 mph and no flagger is used (see Figure 4-6). The free flow speed is assumed to be 5 mph above the speed limit. More details on how this curve was developed was explained by Benekohal et al. (2010).

In this study, the roadway is discretized using different section length of L ranging from 50 ft to 450 ft. Parts 1 and 2 may or may not be divisible to some of these section lengths. If they are divisible, then all section lengths are equal to L. If Part 1 is not divisible, the length of Section 0 is L + mod (1000, L) and the length of the other sections will be equal to L. Similarly, if Part 2 is not divisible, the length of the Section N will be L + mod (5280, L) and the length of the other sections in Part 2 will be equal to L.

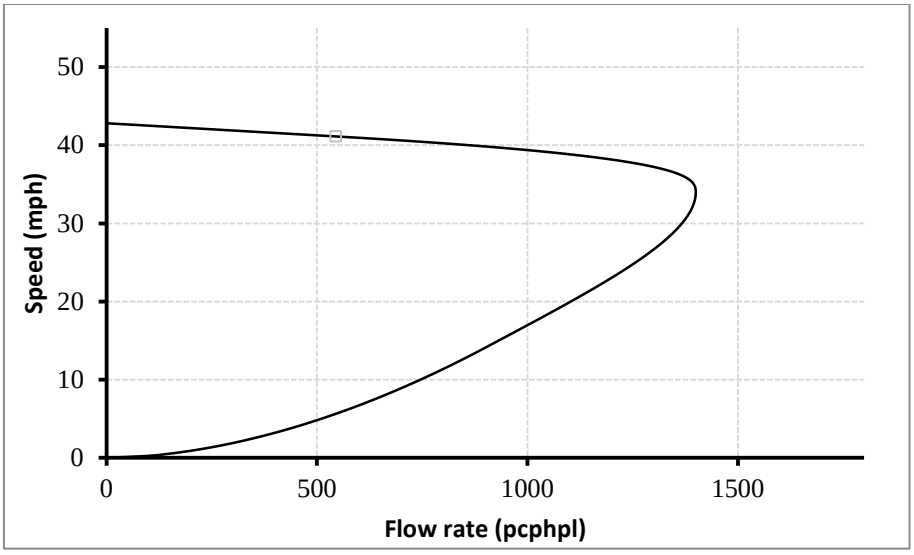


Figure 4-5: Speed-flow curve for part 1: speed limit of 45 mph and flagger

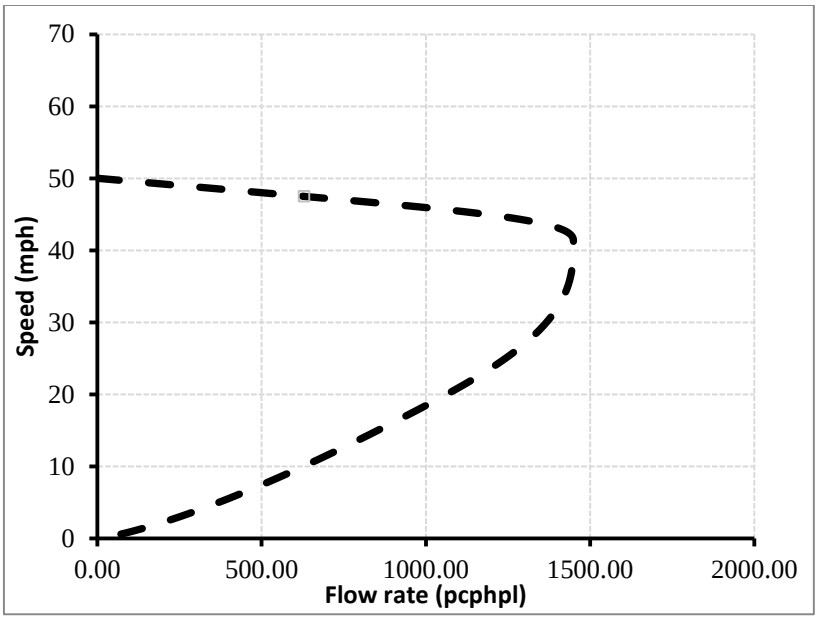


Figure 4-6: Speed-flow curve for part 2: speed limit of 45 mph

4.3 Formulation

Variables, parameters and sets are defined in chapter 3. The equations are formulated in this section.

Eq. 1) Traffic density is estimated using the conservation law:

$$D_{i,t+1} * L = D_{i,t} * L + \left(F_{i+1,t} * \lambda_{i+1} - F_{i,t} * \lambda_{i}\right) * \Delta t \qquad \forall i \in I, \forall t \in T \qquad 4\text{-}3$$

Eq. 2) Speed dynamics is determined based on the Payne model (Payne 1971) which is a "second order" hydrodynamics model. As suggested by Papageorgiou (1998), the computed speed should be capped by zero since there might be possibilities that the Payne model computes a negative speed.

$$V_{i,t+1} = \text{Max}\left\{0, V_{i,t} + \frac{\Delta t}{L} V_{i,t}\left(V_{i+1,t} - V_{i,t}\right) + \frac{\Delta t}{\tau}\left(V(D_{i,t}) - V_{i,t}\right) - \frac{\Delta t * \vartheta}{\tau * L}\left(\frac{D_{i-1,t} - D_{i,t}}{D_{i,t}}\right)\right\} \quad \forall i \in I, \forall t \in T \qquad 4\text{-}4$$

Eq. 3: The speed-flow curves are converted to speed-density relationships as follow. More details on conversion can be found in Chapter 5.

$$V\left(D_{i,t}\right) = \begin{cases} \dfrac{42.81}{1 + 0.0031 * D_{i,t}} & \text{if } D_{i,t} \le 13.26 \\ \dfrac{62.85}{1 + 0.01863 * D_{i,t}} & \text{if } 13.26 \le D_{i,t} \le D_{cr} = 41.18 \\ \left(\dfrac{D_{i,t}}{250}\right)^{\frac{1}{-0.532}} & \text{if } D_{cr} \le D_{i,t} \end{cases} \qquad \forall i \in I_1, \forall t \in T \qquad 4\text{-}5$$

$$V(D_{i,t}) = \begin{cases} \frac{50}{1+0.004*D_{i,t}} & \text{if } D_{i,t} \leq 13.26 \\ \frac{53.54}{1+0.00865*D_{i,t}} & \text{if } 13.26 \leq D_{i,t} \leq D_{cr} = 36.04 \\ \left(\frac{D_{i,t}}{250}\right)^{\frac{1}{-0.505}} & \text{if } D_{cr} \leq D_{i,t} \end{cases} \quad \forall i \in I_2, \forall t \in T \qquad 4\text{-}6$$

Eq. 4) Exit flow rate is defined as:

$$F_{i,t+1} = D_{i,t+1} * V_{i,t+1} \quad \forall i \in I, \forall t \in T \qquad 4\text{-}7$$

4.4 Calibration method

The parameters are found such that the second order model returns queue length and speed estimates that are similar to field data. The calibration methodology that will be proposed in this section has two phases. Phase 1) Search Trajectory Recognition (STR), and 2) Local Search (LS). The Phase 1 determines a search direction and a search boundary that are used for Local Search in Phase 2. To determine the search direction, the study looks at goodness of fit measures for only speed data to recognize how the quality of fit changes in certain search directions, and to find a search direction that facilitates finding the parameters. Also, the study determines boundaries of the search area that most likely includes the appropriate parameters. Then in Phase 2, a local search is performed to fit both queue length and speed data.

4.4.1 Phase 1: Search Trajectory Recognition (STR)

This Phase has two stages:

Stage 1) Direction Recognition

Stage 2) Boundary Recognition

4.4.1.1 Stage 1: Direction recognition

The quality of fit is evaluated for the following directions:

Direction 1) τ is fixed and ϑ changes

Direction 2) ϑ is fixed and τ changes

Direction 3) both ϑ and τ change, but $R = \frac{\tau}{\vartheta}$ is fixed

Our experiment is designed such that enough evidences are found to understand how average and standard deviation of error in speed estimates change for each direction. The direction evaluation will be performed over a range of τ and ϑ shown in the Table 4-1. The

ranges of the parameters are determined based on the reported values in the literature. The reported values for the anticipation coefficient range from 1.5 $\frac{\text{mile}^2}{\text{hr}}$ (Payne 1976) to 23.4 $\frac{\text{mile}^2}{\text{hr}}$ (Kotsialos et al. (2002)). This range is extended slightly in this study and the anticipation coefficients are between 0.5 $\frac{\text{mile}^2}{\text{hr}}$ and 25 $\frac{\text{mile}^2}{\text{hr}}$with increment of 2.5 (except the first increment which is 2.0). The minimum relaxation time value reported in calibration studies is 1.8 sec. (Traffic Flow theory Monograph, Revised Chapter 5, 2002). Thus the relaxation time ranges from 2 seconds to 98 seconds with increment of 4 seconds. The results did not show that these ranges should be extended.

This study does not examine all the possible pairs. It only selects the pairs that represent R($=\frac{\tau}{\vartheta}$) values of 2, 4, 10, and 20 ($\frac{\text{hr}^2}{3600\ \text{mile}^2}$) as highlighted in Table 4-1. The values of R for the cells with the same color are either equal or very close. It should be noted that the value of $\frac{\tau}{\vartheta}$ may not be identical for all pairs of the same color since the ranges of ϑ and τ have been discretized. The cells with R values of 2, 4, 10, and 20 are in red, yellow, green, and blue, respectively.

The computations were performed for L=150 ft and Δt =1.5 seconds. These selections of L and Δt will not cause instability in the computations since the ratio of $\frac{L}{\Delta t}$ is 68 mph which is more than the maximum speed of 50 mph that is possible in the computations. Using Equations 4-1 to 4-4, speed, density and flow rate were computed for each Δt -second interval and each section. The speed values for the section 1 were aggregated over 1-minute intervals and compared with 1-min average speed computed based on field data. The error e_m for each minute is defined as

$$e_m = \frac{V_{1,m} - V_{f,m}}{\sum m} \qquad m \in \{1, \dots, 48\} \qquad 4\text{-}8$$

Where,

m: Minute index.

$V_{1,m}$: Average Speed for section 1 and minute m computed based on the second order model.

$V_{f,m}$: Average Speed for minute m computed based on field data.

Each highlighted cell of the Table 4-1 displays average error and standard deviation of error (inside the parenthesis). A negative (positive) average error value means that the model underestimates (overestimates) speed. The next section discusses the observed trends.

Table 4-1: Average error (standard deviation) of speed estimates for different directions

Relaxation time (τ) in sec	Anticipation coefficient, ϑ, in $\frac{mile^2}{hr}$											
		0.5	2.5	5	7.5	10	12.5	15	17.5	20	22.5	25
	2	-0.71 (5.4)										
	6	8.04 (9.5)	-8.41 (5.1)									
	10	8.35 (9.9)	-4.4 (6.2)	-9.53 (6.6)								
	14				-9.6 (6.7)							
	18					-9.68 (6.9)						
	22			-9.4 (6.6)								
	26		4.04 (9.4)				-10.1 (7.3)					
	30				-11.1 (6.4)			-9.95 (7.2)				
	34								-9.9 (7.2)			
	38									-9.84 (7.1)		
	42					-11.5 (6.9)						
	46										-10.1 (7.4)	
	50		13.5 (7.1)	-0.991 (11.7)			-12.3 (7.6)					-10 (7.3)
	54											
	58											
	62							-12.4 (7.6)				
	66											
	70								-12.7 (7.9)			
	74				-3.15 (12.1)							
	78											
	82									-12.7 (7.9)		
	86											
	90										-12.9 (8)	
	94											
	98			13.7 (7.2)		-4.35 (11.9)						-12.9 (8)

Color code
R=2
R=4
R=10
R=20

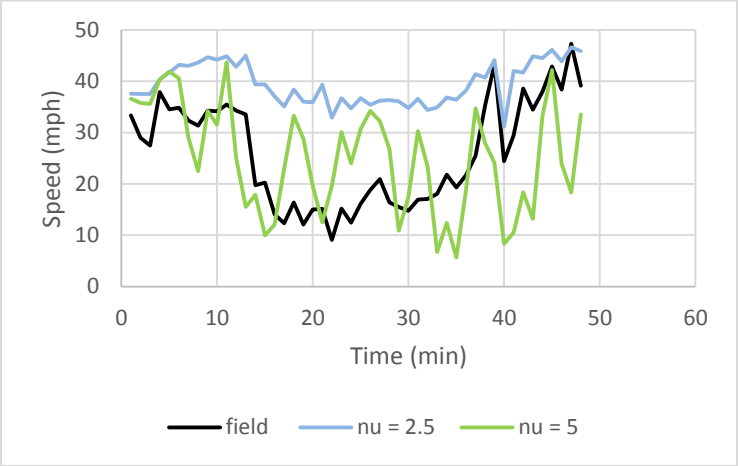
4.4.1.1.1 Observed trends for direction recognition stage

Trends for Direction 1 (τ is fixed and ϑ changes): As shown in Table 4-1overall, the model overestimates average speed for low values of ϑ and the amount of overestimation reduces as ϑ increases. To demonstrate this trend, Figure 4-7 a and b show the estimated speeds when τ

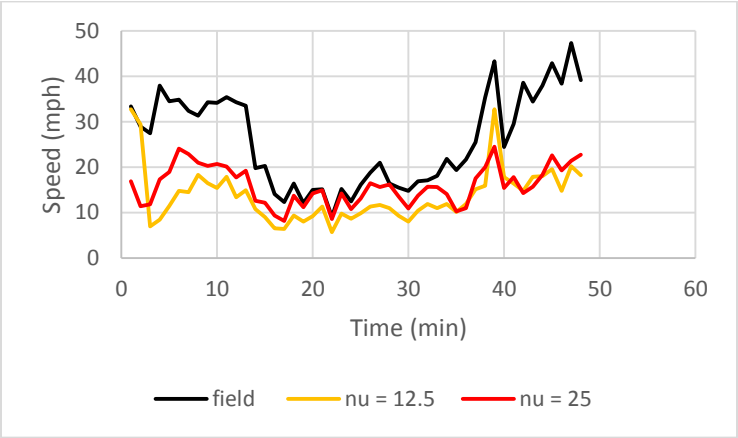
is equal to 50 sec and ϑchanges as highlighted in Table 4-1. When ϑ is equal to 2.5, the ratio of $\frac{\vartheta}{\tau}$ is not large enough to create speed reductions due to density change imposed by boundary conditions; thus speed did not drop sufficiently to match field data and it is above 30 mph. When ϑ is increased from 2.5 to 5, the model becomes more sensitive to density change and average error in speed becomes -0.99 mph which is very low; however speed considerably fluctuates as shown in Figure 4-7a and standard deviation of error in speed estimates is 11.7 mph which is high. In this case, the model overestimates speed for 26 time intervals and underestimates speed for 22 intervals, while error in speed ranges from -16.9 mph to 29 mph.

When ϑ is further increased from 5 to 12.5 and then to 25, the speed is underestimated. This is because $\frac{\vartheta}{\tau}$ is large and as a result, the model is oversensitive to density changes. This trend implies that probably a ϑ between 2.5 and 12.5 will return admissible match between model estimates and field data.

Trends for Direction 2 (ϑ is fixed and τ changes): In Table 4-1, the model overestimates speed as τ increases. To demonstrate the trend, Figure 4-8 a and b display speed estimates when ϑ is equal to 2.5 and τ varies as highlighted in Table 4-1. When τ is equal to 6 sec, the speed is underestimated for all intervals. This is because the ratio of $\frac{\vartheta}{\tau}$ is large and the model is too sensitive to density increases. If τ is increased from 6 sec to 10 sec, this sensitivity becomes less and for some intervals, especially after minute 40, the model can show a trend closer to field data. When τ is further increased to 26 sec, the model overestimates average speed by 4.04 mph. For τ value of 50 sec., the model overestimates average speed by 13.5 mph.



Part a

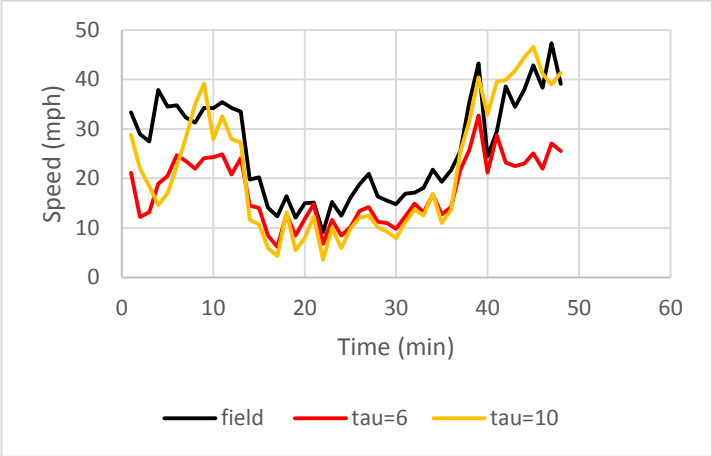


Part b

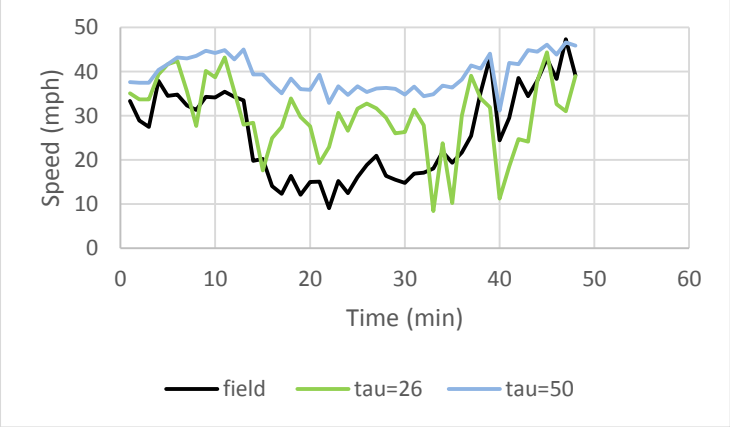
Figure 4-7: Direction 1 evaluation: Speed estimates for ϑ values of a) 2.5 and 5 and b) 12.5 and 25

The observed trend implies that appropriate values of τ are between 10 and 26 when ϑ is 2.5. Thus one may find an appropriate value of τ by applying search methods, such as bisection method, when ϑ is fixed and τ varies from 10 to 26; however this direction is less preferable than direction 1 because changing τ, affects two terms that may have interaction: 1) density gradient term $\left(\frac{T*\vartheta}{L*\tau}\left(\frac{D_{i-1,t}-D_{i,t}}{D_{i,t}+\kappa}\right)\right)$ and 2) relaxation term$\left(\frac{T}{\tau}\left(V(D_{i,t})-V_{i,t}\right)\right)$. Alternatively in direction one, only ϑ changes and influences the gradient term. Moreover, fixing ϑ and changing τ, may return a relaxation time that is not physically justifiable. In particular, relaxation time is a function of drivers' reaction time and cannot take any large values. To avoid

large τ values, it is better to use direction 1 since τ can be set up to a value that is physically justifiable and then the search algorithm finds the corresponding value of ϑ.



Part b



Part b

Figure 4-8: Direction 2 evaluation: Speed estimates for τ values of a) 6 and 10, and b) 26, and 50

Trends for Direction 3 (Both ϑ and τ change, but $R = \frac{\tau}{\vartheta}$ is fixed): In Table 4-1 for τ values of greater than or equal to 10 seconds, the estimates are mainly characterized by a value of R.

For R value of 2, the model underestimates speed. Figure 4-9a shows speed estimates when R is equal to 2 for two pairs of (τ = 10, ϑ = 5) and (τ = 50, ϑ = 25). Although there is a

considerable difference between τ values in the pairs, there is no significant difference between the estimates.

For R value of 4, the model underestimates speed for most of τ values, especially τ values of greater than 10 seconds. As shown in Figure 4-9b, the model underestimates speed when τ is equal to 22 and 50; However for τ value of 6 the model estimates are close to field data with average error of -0.71. The trend observed for R values of 2 and 4 suggests the hypothesis that for a given R, the model is not considerably sensitive to changes in τ values. The next stage, Boundary Recognition, will closely explore this hypothesis and determine the τ values beyond which the model is not sensitive to τ .

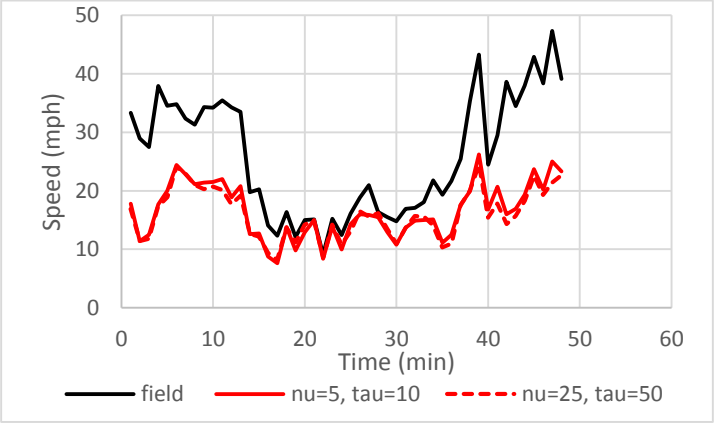
For R value of 10, although the model sometimes (e.g. τ = 6) overestimates and sometime (e.g. τ = 74) underestimates speed, all these estimates are characterized by high speed fluctuations and high standard deviation of error as shown in Figure 4-10b.

For R value of 20, the model overestimates speed. Figure 4-10c shows the estimates when R≈20 for different pairs of (ϑ = 0.5, τ = 10) and (ϑ = 5, τ = 98). Both pairs overestimate speed. This might be because R is relatively large in the density gradient term and cannot reproduce speed drop and queuing conditions.

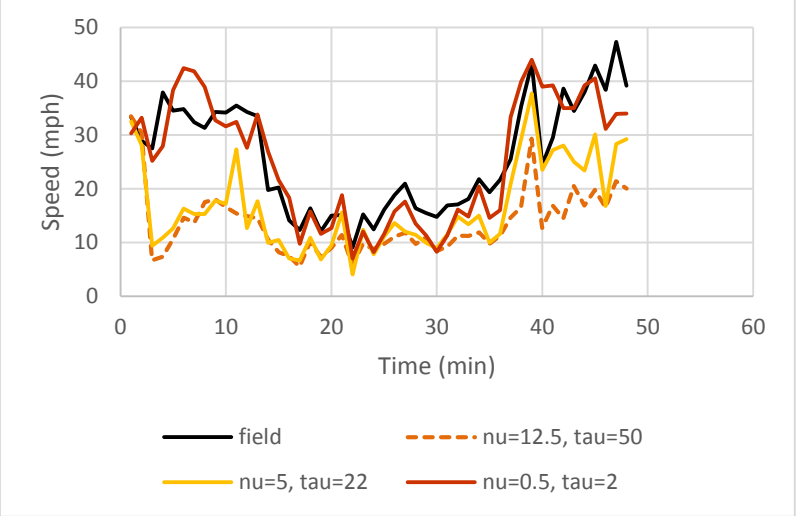
The observations show that changing ϑ and τ values in the direction 3, may not considerably influence the goodness of fit. For instance, R = 2 and R = 4 mostly underestimate speed, and R = 10 overestimates speed. Thus the search direction 3, is not good because it can reproduce similar trends. However this information is not useless and can determine boundaries of the search. For instance in this case, probably the proper value of R is between 4 and 10. Thus, the search boundaries can be determined based on R values in Stage 2.

Another important observation is that there might be different pairs which can return similar goodness of fit measures. These pairs may have either an unrealistically large relaxation time or a justifiable relaxation time. For instance, for τ =10 and ϑ=5, the average error is -9.53 and for τ =38 and ϑ=20 the average error is -9.84. This could be considered as an indication that there might be different pairs which return similar and/or acceptable goodness of fit measures; however because of non-convexity and perhaps presence of many local minima, an optimization algorithm may converge to a pair which has an unreasonable value of relaxation time. This suggests that search or optimization algorithms should be performed over a direction which does

not have many local minima or maxima. The examination of different directions showed that using direction 1 would reduce the likelihood of stopping at local minima.

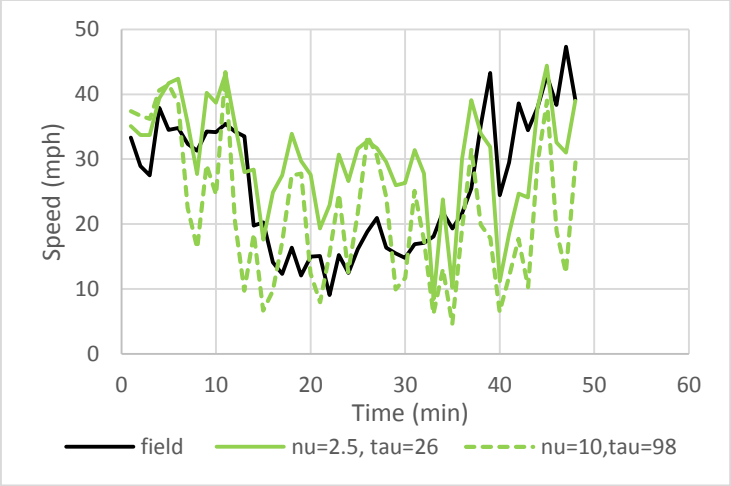


Part a

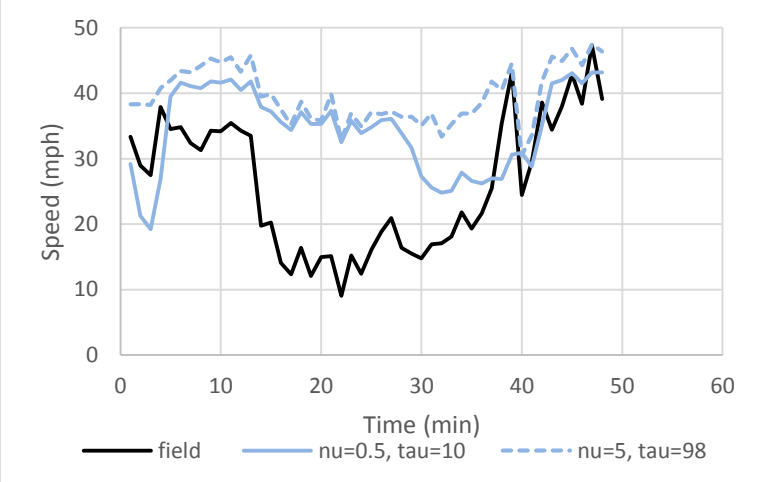


Part b

Figure 4-9: Direction 3 evaluation: Speed estimates for R values of a) 2 , and b) 4



Part a



Part b

Figure 4-10: Direction 3 evaluation: Speed estimates for R values of a) 10, and b) 20

Based on these observations and comparison of the directions, the following decisions are made:

- Direction 1 (i.e. τ is fixed and ϑ changes) is chosen for search direction. This is because τ can be fixed to a reasonable value that can be justified based on the reaction time of drivers. Moreover, it seems that increasing the value of ϑ can reduce speed in the queue and increase the propagation rate; as a result this direction will help to avoid trapping in local minima.
- The search range for ϑ is limited based on the values of R. This is because the value of R is the main parameter in the density gradient term creating queue. Beyond some R values

(in this case 10) the model overestimates speed, and below some R values (in this case 4) the model underestimates speed in the queue.

The Figure 4-11 shows the schematics of search direction and boundaries. The next section will determine the boundaries of the search area.

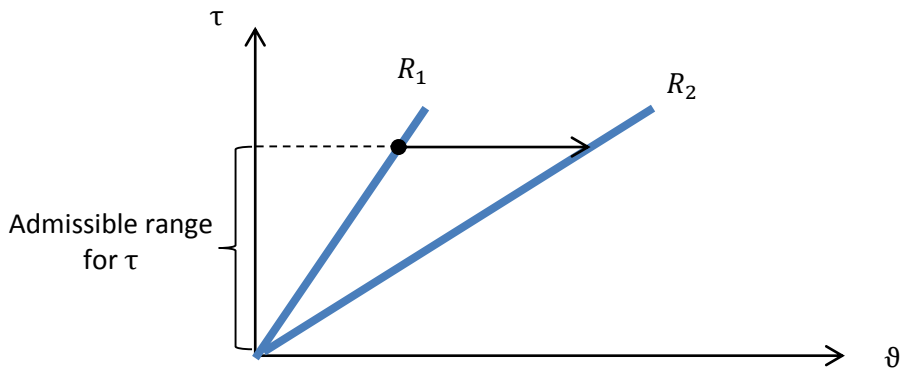


Figure 4-11: Schematic representation of search direction

4.4.1.2 Stage 2: Boundary recognition:

The values of τ and R are determined as below:

- The lower bound of the τ is 2 second which is around the reaction time of a driver. Upper limit of τ is the maximum amount of time that a section needs to reach to a steady state condition. This could happen when maximum number of vehicles that are present in a section and depart a stopped queue with a headway of 2 seconds (saturation flow rate of 1800 pcphpl). Thus

  Upper limit of τ = the number of vehicles in stopped queue* 2 sec

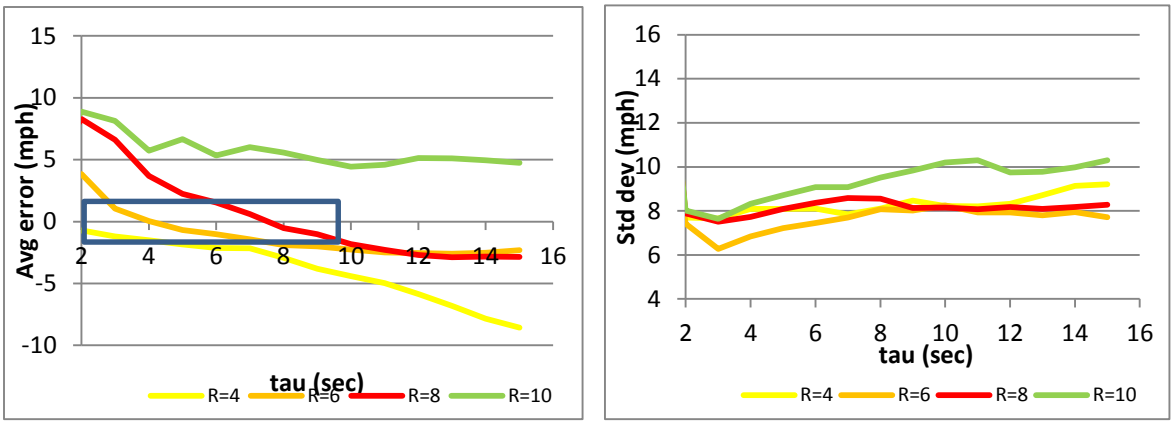
  The number of vehicles in stopped queue is computed as the section length divided by 20 ft which is the spacing between vehicles in stopped queue. Thus, the upper limit can be simply computed as:

  Upper limit of τ = section length in ft/10

  Based on the equation above, the upper limit of τ is 15 sec for section length of 150 ft.
- R should be limited between 4 and 10. Beyond 10, the model overestimates speed and below 4, the model underestimates speed.

One can use a local search algorithm at this point to match both speed and queue length data. Alternatively, the study explores the model to further limit the search area. Figure 4-12a and Figure 4-12b show the average error and standard deviation of error in speed estimates for

the search area detected above. The study further limits search area to R values that return an average error between -2.5 mph and 2.5 mph. In addition, the upper part of the search range for τ is limited by τ of 9 second because beyond this relaxation time either the change in average error in speed is not significant (e.g. R=8) or the average error is larger than ±2.5 (e.g. R=4). The box in Figure 4-12a indicates the search area which will be used to fit queue length data. Figure 4-12b shows the variation of standard deviation. For the selected range of τ, the variation in standard deviation is not significant thus this stage does not further shrink the box indicated based on average error.



Part a Part b

Figure 4-12: Average error (Part a) and standard deviation (Part b) for L=150 ft, and $\frac{L}{\Delta t} = 68$ mph

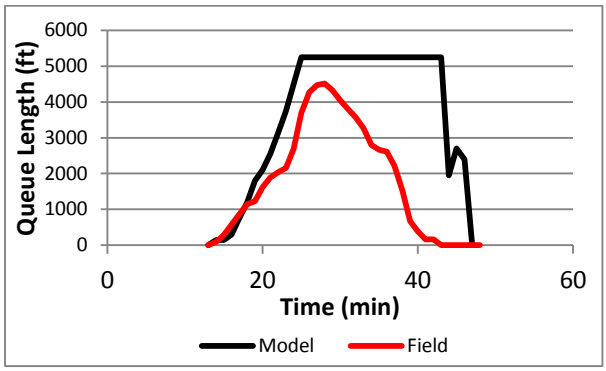
Establishing the box in Figure 4-12a is based on the assumption that beyond relaxation time of 9 sec the queue length estimate are not sensitive to τ values. The validity of this assumption will be explored once the parameters are calibrated in Phase 2. We will compare the fitted queue length for τ values inside and outside the box to explore the validity of the assumption. This will determine if there are significant changes in queue length estimates when τ values outside the box are used.

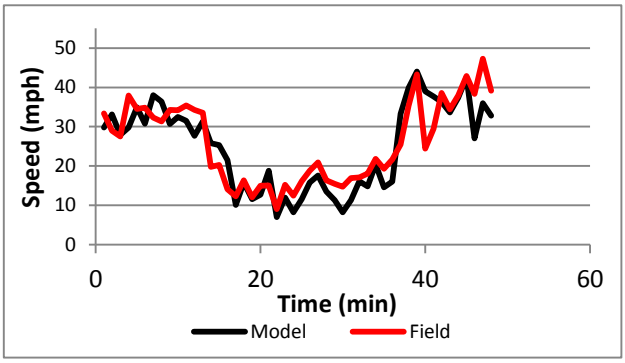
4.4.2 Phase 2: Local search

For each value of τ, the search is performed to find ϑ values with accuracy of 0.1 returning an average discrepancy of less than 120 ft for the average queue length. This error is the same as the accuracy of queue lengths measured in the field. In the computations, a section is declared to be in full queuing condition when the speed drops below 20 mph this is indicated

based on the field data in which speed varies between 9.1 mph and 21.7 mph when camera captured queued vehicles (see Figure 4-2 from minute 13 to minute 36).

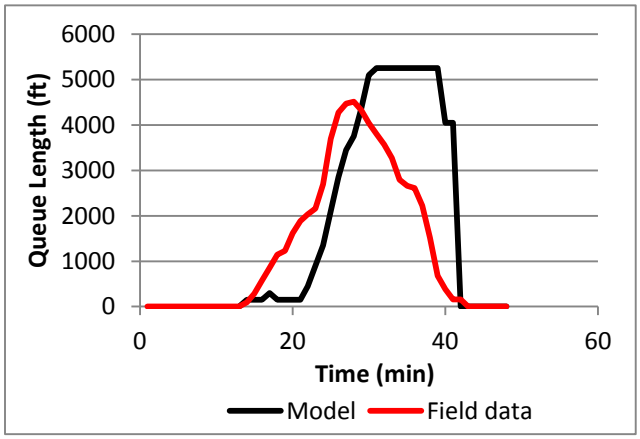
The search results showed that for τ values of less than 7 seconds there is not a good fit. In particular, when average error in speed is between ±2.5 mph, the error in average queue length is more than 120 ft. As shown in Figure 4-13a, when τ = 2 sec and ϑ = 0.5 (R=4) the queue length is overestimated while speed estimates are close to field data with average error of -1.13 (See Figure 4-13b). To reduce the amount of queue length overestimation, one can reduce the value of ϑ. If ϑ is reduced to 1/3 (R=6) (See Figure 4-14), still queue is overestimated while the average error in speed becomes 3.72 mph which is more than 2.5 mph. If ϑ is further reduced, the average error in speed becomes much higher than 2.5 mph; thus there is not a pair of τ and ϑ that returns good fit. There is a similar condition for other values of τ less than 7 seconds.

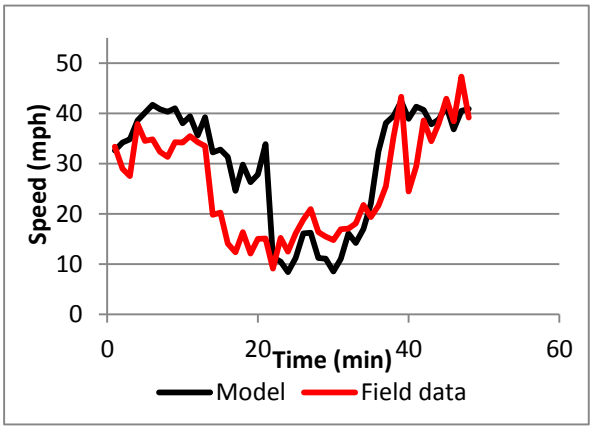




Part a Part b

Figure 4-13: Results for τ =2 sec and ϑ =0.5: a) Queue length b) Speed

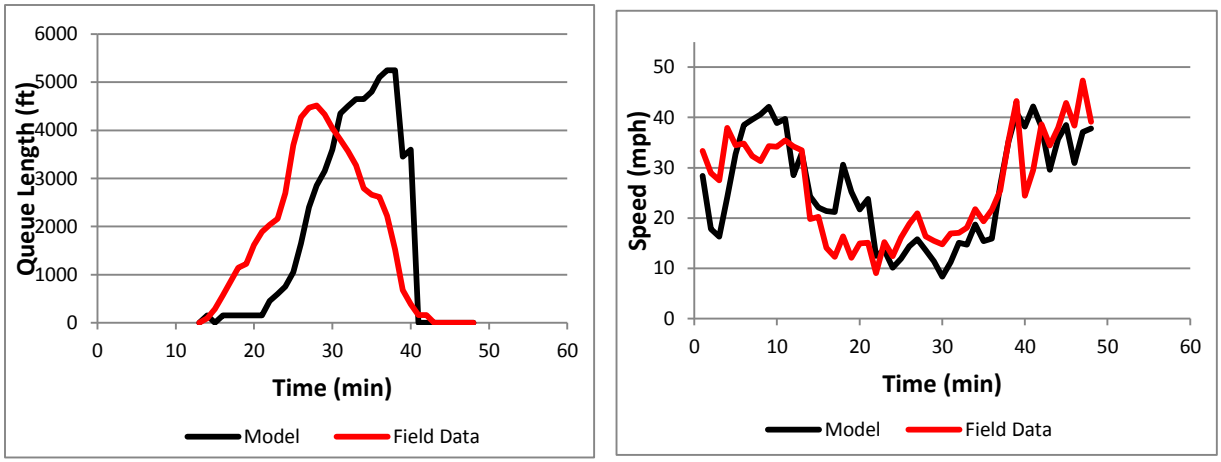




Part a Part b

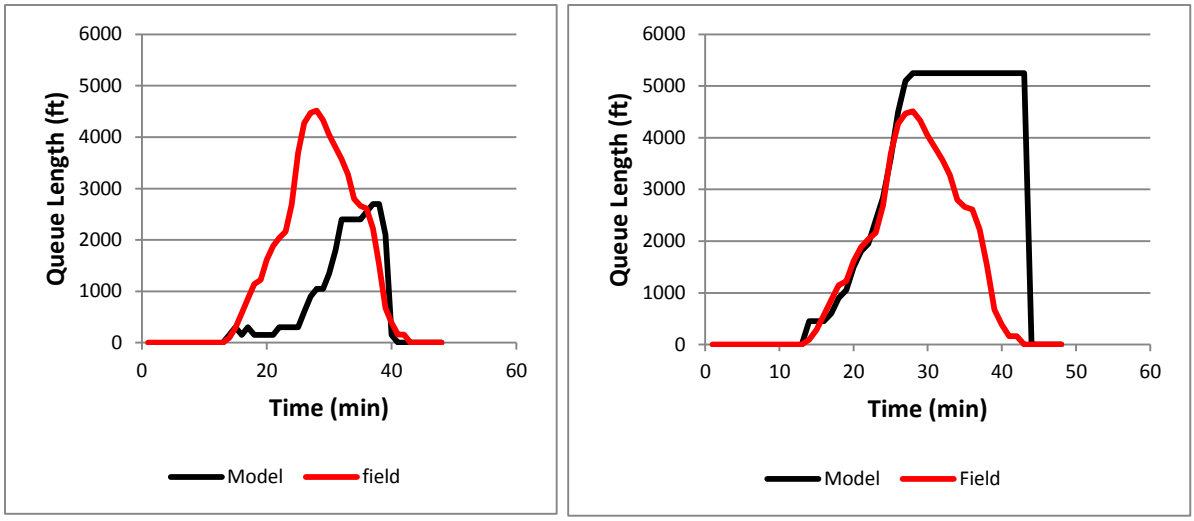
Figure 4-14: Results for τ =2 sec and ϑ =1/3 a) Queue length b) Speed

For τ values of 7 seconds or higher, local search could find ϑs returning acceptable average error values. Figure 4-15 displays queue length and speed estimates for τ =7 sec and ϑ = 0.92 returning average error values of -0.096 mph and -18.7 ft for speed and queue length, respectively. Increasing (decreasing) ϑ will increase (reduce) queue length. For example as shown in Figure 4-16a , if ϑ is reduced from 0.92 to 0.85, the average queue length will be reduced from 2037 ft to 944 ft which is corresponding to 54% of queue length reduction. If ϑ is increased from 0.92 to 1, queue length will be overestimated as displayed in Figure 4-16 b; the rate of propagation is more similar to that in the field, but the rate of shrinkage is not fast enough to replicate queue shrinkage in the field. As a result, back of queue reaches the very upstream section and queue covers the whole roadway (5280 ft) for 15 minutes. This observation suggests modifying the second order model to have two different ϑ values: one for propagations and the other one for shrinkage. The ϑ value for shrinkage should be larger than that for propagation to trigger queue shrinkage earlier and avoid overestimation in queue length. Modifying the model using two ϑ values will be investigated later in this chapter.



Part a Part b

Figure 4-15: Results for τ =7 sec and ϑ = 0.92 a) Queue length b) Speed



Part a Part b

Figure 4-16: Queue length for τ =7 sec and a) ϑ = 0.85, b) ϑ =1

The parameters were found only for relaxation time values of 7, 8 and 9 seconds as indicated by the box shown in Figure 4-12a. The values of the parameters are reported in Table 4-2. The relaxation time value did not have significant influence on the shape of queue length. Figure 4-17 shows the fitted queue lengths for the relaxation times of 7 and 9 that are low and high ends indicated by the box. The estimated queue lengths are close with maximum discrepancy of 300 ft and average discrepancy of 77 ft.

The local search was performed for the τ values inside the box assuming that τ values greater than 9 seconds do not substantially change the quality of fit for queue length. This assumption is checked by comparing quality of fit for τ = 15 seconds, which is outside the box, versus τ values of 7 and 9 seconds, which are inside the box. For τ = 15 seconds, the ϑ value of 2.16 returns acceptable average error values of -2.43 mph for speed and -68 ft for queue length. As shown in Figure 4-17, when τ is 15 seconds, the propagation starts earlier than when τ is 7 or 9. This causes queue length to be slightly longer before minute 26 when τ is 15 seconds. After minute 26, the estimated queue lengths are comparable for the selected τ values. These comparisons, along with the trends in Figure 4-12, show that τ values greater than 9 seconds would not influence the quality of queue length estimates and one may eliminate those values from the search area.

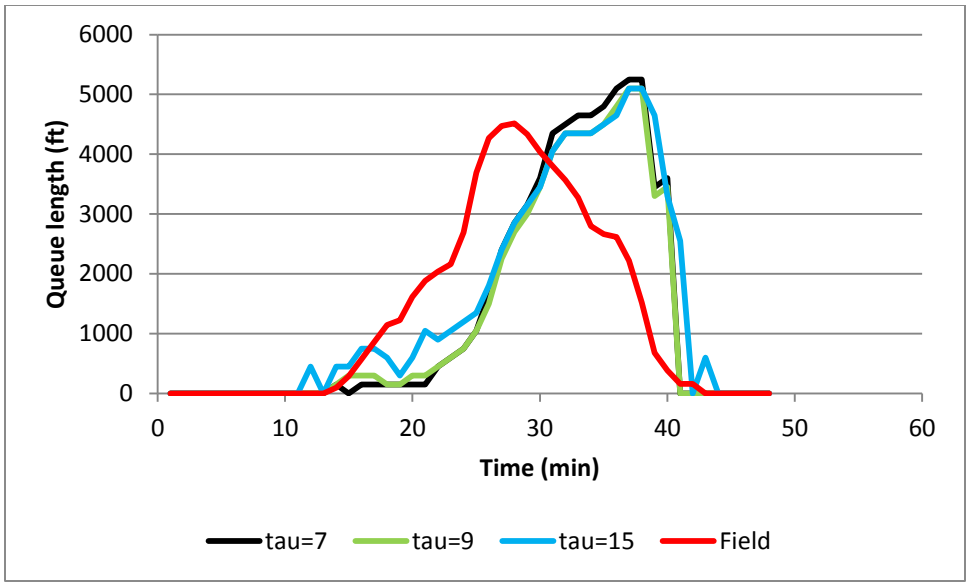


Figure 4-17: Fitted queue length for different values of relaxation time (τ)

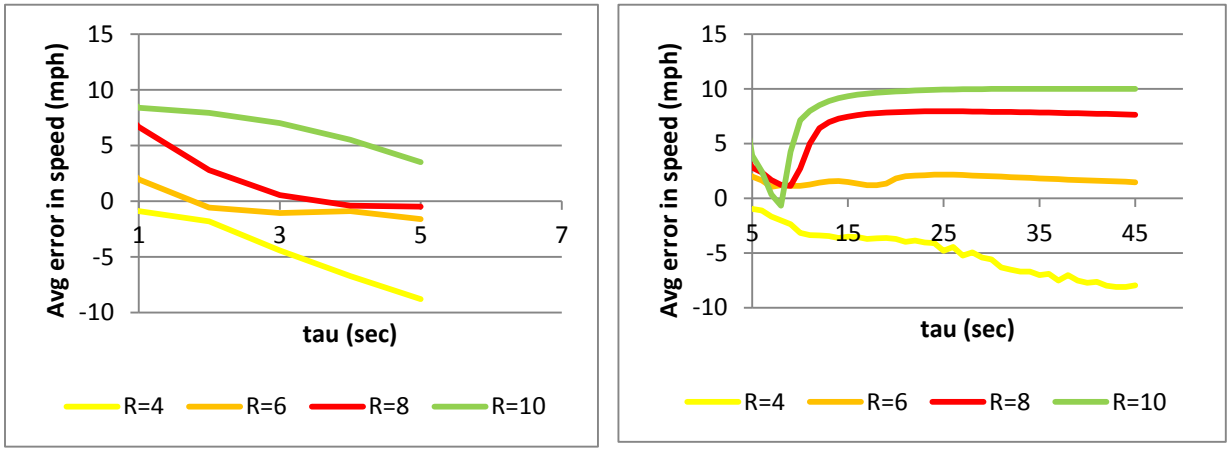
Table 4-2: Goodness of fit measures for the calibrated parameters when L=150 ft and $\frac{L}{\Delta t}$=68 mph

L (ft_)	Δt (sec)	τ (sec)	ϑ	Average error in speed (mph)	Standard deviation of error in speed (mph)	Average error in queue length (ft)
150	1.5	7	0.92	-0.096	8.19	-18.7
		8	1.02	-0.693	8.31	111.3
		9	1.14	-1.52	8.05	-86.5
			1.13	-1.41	8.15	102.2

4.5 **Sensitivity analysis** for **section length (L)**

The analysis are conducted for different values of section lengths: 50 ft, 100 ft, 150 ft, 300 ft, 450 ft while $\frac{L}{\Delta t}$ is equal to 68 mph for all the lengths. Figure 4-18 shows how average error in speed varies with respect to R for the shortest and longest section lengths. The trend is similar to that found for L=150 ft in Figure 4-12a. Overall, as R increases, the speed is overestimated. As a result, the search area is limited to R values between 4 and 10 that return an average error between -2.5 mph and +2.5 mph. Also, the trend shows that after some τ value, average error and standard deviation values are not sensitive to the relaxation time. This threshold for τ varies

based on section length as shown in Table 4-3. Thus, the search area is further limited to τ values less than these thresholds.



Part a Part b

Figure 4-18: Average speed for different R values: a) Section length = 50 ft, b) Section length = 450 ft

Table 4-3: Upper end of relaxation time (τ) for different values of section lengths

Section length (ft)	**Upper end of search range for τ (sec)**
50	5
100	7
150	9
300	11
450	13

The search was performed to find the parameters. Similar to L=150, there was not good fit when τ is less than 7 seconds because either the average error in speed was more than 2.5 mph or average error in queue length was more than 120 ft. This means that there is not a good fit for section lengths of 50 ft and 100 ft. Beyond the numerical results, conceptually one can argue that section lengths of 50 ft and 100 ft are not as good as the longer section lengths. A section length should be long enough to encompass at least one vehicle at each time interval; otherwise density fluctuates between zero and a high number and this could be a cause of instability in the results. For instance, section length of 50ft is corresponding to density of 105.6 veh/mi/ln =(5280/50) and may be an acceptable choice when density in the queue is greater

than105 veh/mi/ln; however the computed density in the front of queue varies between 40 and 93 pc/mi/ln as shown in Figure 4-19. Section length of 150 ft is corresponding to density of 35.2 pc/mi/ln which is lower than density values in the queue. As a result, selection of 150 ft or higher for the section length can include at least one vehicle in the section when there is queue in the field. Section lengths of 300 ft and 450 ft could be more preferable since they are corresponding to density values of 17.6 pc/mi/ln and 11.7 pc/mi/ln and can include one vehicle in the section when traffic is in undersaturated conditions.

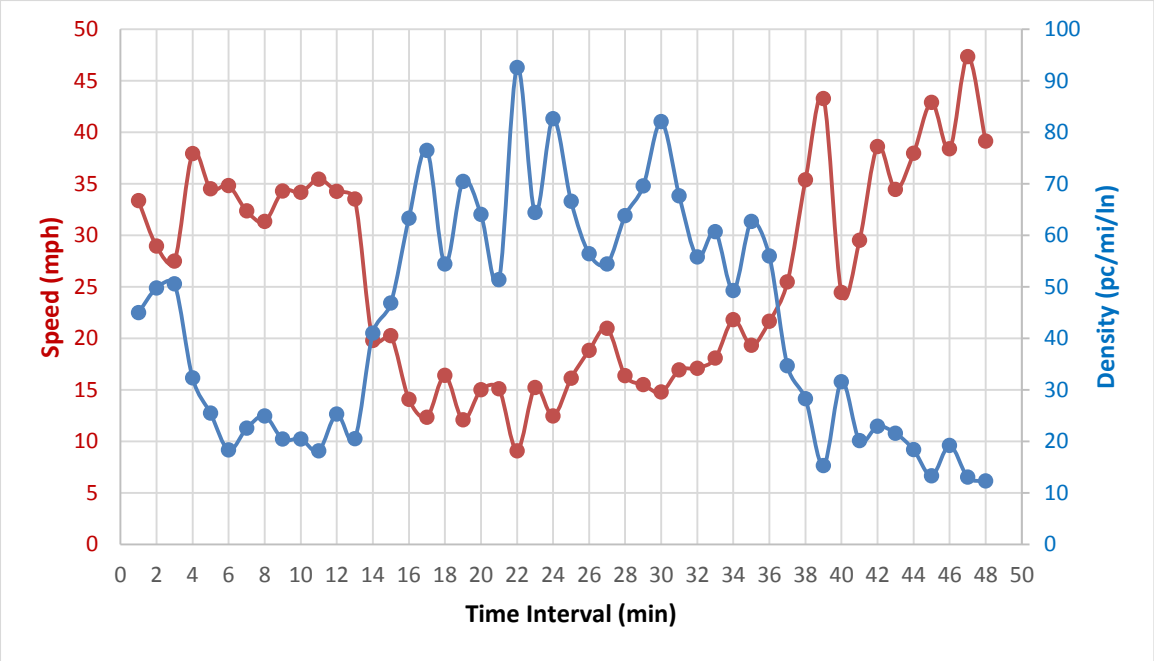


Figure 4-19: Speed data collected from field and computed densities

The calibrated parameters are shown in Figure 4-20 for sections lengths of 150 ft to 450 ft. There is an increasing trend between anticipation coefficient and relaxation time. The anticipation coefficient varies between 0.92 and 2.31, and for a given relaxation time, the anticipation coefficient increases as section length increases.

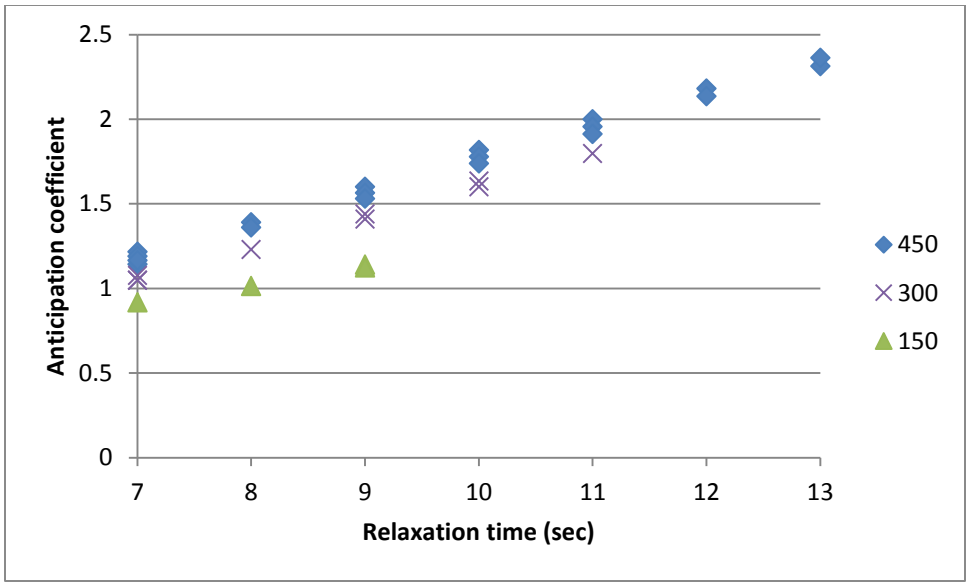


Figure 4-20: Calibrated parameters for different section lengths

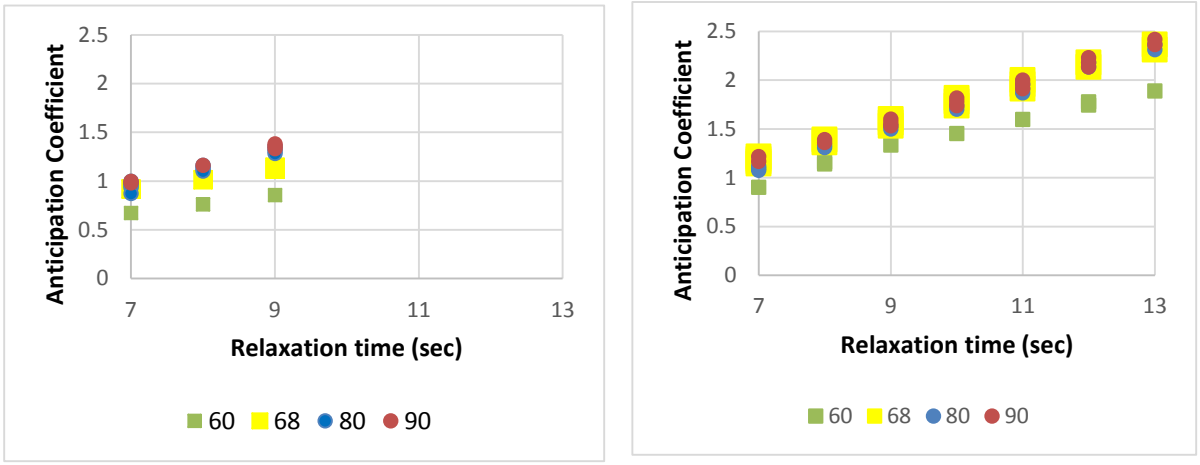
4.6 Sensitivity analysis for $\frac{L}{\Delta t}$

Sensitivity analysis was also performed for $\frac{L}{\Delta t}$ of 60, 68, 80, and 90 mph. Figure 4-21 shows the trend between the calibrated parameters for L=150 ft and L=450 ft. The value of anticipation coefficient increases as $\frac{L}{\Delta t}$ increases and there is a linear trend between ϑ and τ. A linear regression is developed with the general form of:

$$\vartheta = a + b\,\tau \qquad 4\text{-}9$$

Where, a and b are the regression coefficients.

For L=150, lines were fitted for each $\frac{L}{\Delta t}$ and the R-square values and coefficients are reported in The R-square values for all $\frac{L}{\Delta t}$s are greater than 0.95. Figure 4-21b suggests that the data points for $\frac{L}{\Delta t}$ of 80 and 90 may be combined. For the combined data, the linear regression results in R-square of 0.96. Estimation error was computed for the combined data as the difference between the estimated ϑ and the actual ϑ found from the calibration. The average error is equal to zero and error ranges from -0.050 to 0.084.



Part a Part b

Figure 4-21: Calibrated parameters for different values of $\frac{L}{\Delta t}$ a) L = 150 ft, b) L = 450 ft

Similar analysis was conducted for other section lengths and Table 4-4 shows the coefficients. The R-square values are 0.96 or higher indicating a good fit. The p-values shows that intercepts (parameter a) are significant for the combined data sets. For the other data sets, the p-values for intercept are 0.08 or higher, but the study does not make a conclusion that the intercepts are insignificant since the number of data points vary between 3 and 9 which are low. Thus the intercepts are kept in the model for all the data sets. The p-values for the slope (parameter b) is 0.013 or lower meaning that the slope is significant and should be kept in the model.

Table 4-4: Regression results for L=150 and different $\frac{L}{\Delta t}$s

$\frac{\mathbf{L}}{\mathbf{\Delta t}}$	60	68	80	90	80, 90 combined
a (intercept)	0.028	0.156	-0.346	-0.346	-0.358
b	0.092	0.108	0.185	0.185	0.188
R-square	0.99	0.99	0.96	0.99	0.96

Table 4-5: Regression results for different section lengths and $\frac{L}{\Delta t}$s

L (ft)	150			300		450	
Data sets	1	2	3	4	5	6	7
$\frac{L}{\Delta t}$ **(mph)**	60	68	80, and 90 combined	60	68, 80, and 90 combined	60	68, 80, and 90 combined
a (intercept)	0.028	0.156	-0.358	-0.056	-0.292	-0.126	-0.248
b	0.092	0.108	0.188	0.130	0.193	0.157	0.200
R-square	0.99	0.99	0.96	0.99	0.99	0.99	0.99
P-value for a	0.313	0.102	<0.0001	0.148	<0.0001	0.084	<0.0001
P-value for b	0.013	0.004	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Number of data points	3	4	20	5	35	9	53

.

4.7 Behavior of the second order model in the τ- ϑ space

The calibration methodology was implemented for different values of section lengths and time increments and in all these trials the second order model showed similar behavior in the τ - ϑ space. Figure 4-22 summarizes the findings. It was found out that the ratio of $\frac{\tau}{\vartheta}$ has the main influence on the speed discrepancies and results showed that $\frac{\tau}{\vartheta} \geq 10$ overestimates speed and $\frac{\tau}{\vartheta} \leq 4$ underestimates speed. In addition, it was shown that beyond some threshold, shown in Table 4-3 for different section length, the relaxation time does not have a significant influence on the speed estimates. As a result, as shown in Figure 4-22, the search area was limited to that portion of the τ - ϑ space that is likely to return acceptable discrepancy.

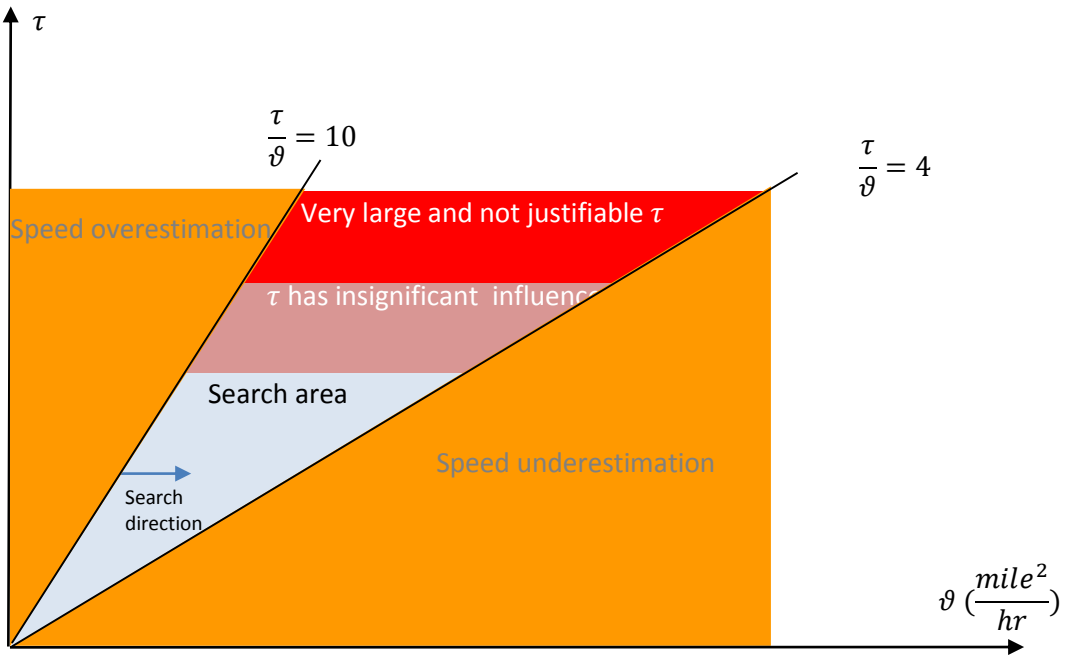


Figure 4-22: Behavior of the second order model in the τ - ϑ space

4.8 Guidelines to select second order model parameter

Based on the results of data fitting and physical meaning of the parameters, the following guidelines are proposed to determine the parameters of the second order model for work zones.

1) Set up:

   $\frac{L}{\Delta t}$ =Free flow speed + θ 4-10

   It is recommended θ to be 10 mph. The results showed that the θ can go up to 40 mph and even it can be half of a free flow speed but a higher value of θ may not be desirable as it causes smaller Δt and more computations. If $\frac{L}{\Delta t}$ is less than free flow speed, instability in the results may occur.

2) The study recommends using section lengths (L) of 300 ft and 450 ft . Section lengths of 50 ft and 100 ft are not recommended because they did not return good results and section length of 150 will increase computations by factor 4 and 9 with respect to section lengths of 300 ft and 450 ft, respectively.

3) Find $\Delta \mathbf{t}$ =L/(Free flow speed + $\boldsymbol{\theta}$) and round it down to the nearest multiple of 0.5 sec.

4) The admissible range of τ for L>=150 ft is determined by τ_{min} and τ_{max} as below

   τ_{min} =max{Δt, 7 sec, 0.0225* section length (ft)}

These lower limits are set up based on the following logic:

- τ has to be greater than Δt ,time interval length ; otherwise the resulting speed would be unstable and will not converge to the steady state speed.
- Based on the field data, the fits are good when τ is equal to or greater than 7 seconds.
- A macroscopic interpretation of relaxation time is the time required for a platoon of vehicle to react. A lower bound for the macroscopic representation, is the time required for half of the vehicles in the section to react. Assuming median reaction time of 0.9 seconds that was proposed for surprised drivers (Johansson and Rumar 1971) the lower bound is:

  0.9 sec*max number of vehicles with 20 ft spacing/2,

  Which can be simplified as

  0.0225* section length (ft)

  τ_{max}, the maximum relaxation time, is the τ value beyond which speed does not significantly depend on relaxation time (See Figure 4-12 and Figure 4-18) and are suggested as follows.

$$\tau_{max} = \begin{cases} \tau = 9\,\text{sec} \quad \text{for L} = 150\ \text{ft} \\ \tau = 11\,\text{sec for L} = 300\ \text{ft} \\ \tau = 13\,\text{sec for L} = 450\ \text{ft} \end{cases} \qquad 4\text{-}11$$

5) Select ϑ based on the linear regression Equation 4-9 with coefficients in Table 4-5.

4.9 Modification of second order model

Two modifications are explored to enhance the second order model:

Modification 1: Add a fine tuning parameter to the denominator

Modification 2: Propose two different anticipation coefficients: one for queue propagation and one for queue shrinkage

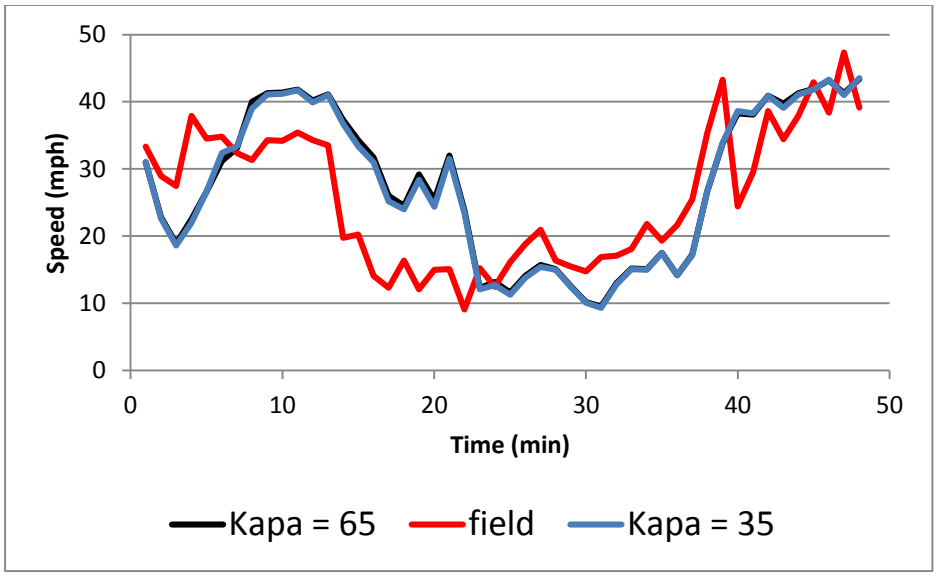
4.9.1 Modification 1: Add a fine tuning parameter to the denominator

When queue propagates and the density of the arriving volume is low the density gradient term, $\frac{D_{i-1,t}-D_{i,t}}{D_{i,t}}$, may become so large and result in unreasonable speed drops. To avoid this condition, Cremer and Papageorgiou (1981) proposed a fine tuning parameter, κ , to be added to the denominator. The Table 4-6 shows different values of κ selected for past studies. The values range from 13 to 40 veh/km/ln (33 to 65 veh/mi/ln).

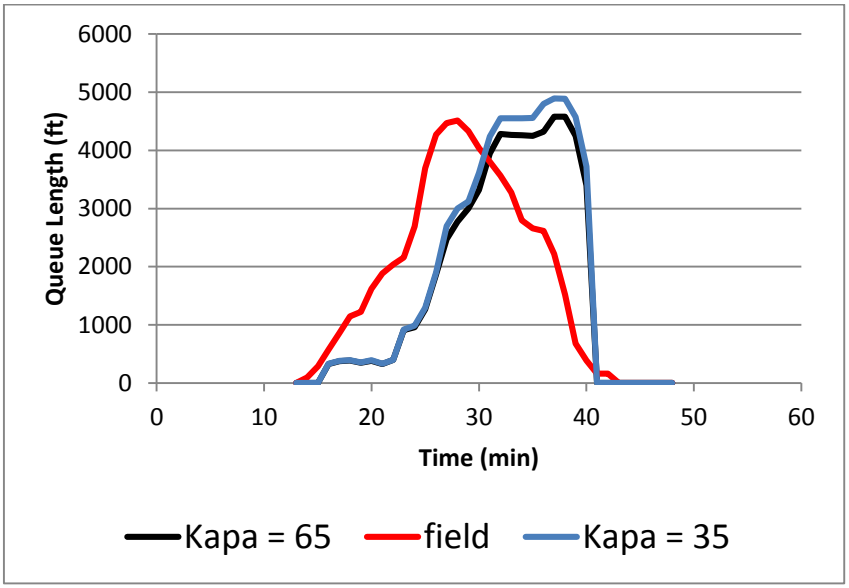
Table 4-6: Fine tuning parameters proposed in the past studies

Year	κ veh/km/ln	Roadway information	Type of volume	Volume (vph) variation over time	Volume (vph) Variation over space	Lane	Volume (vphpl)
Cremer and Papageorgiou (1981)	20	2650 m, no ramp	min-flow rate	1000-3500	_	2	500-1750
Papageorgiou et al. (1990)	40	roughly 1000 m, with one off-ramp and one on-ramp	hourly volume	_	5300-6500	3	1766-2166
Papageorgiou et al. (1990)	13	6 km, 6 off ramp, 5 on ramp	hourly volume	_	data set I:1500-4450	2--3	750-1483
				_	data set II & III:2350-5850	2--3	1175-1950

According to Papageorgiou et al. (1990) any reasonable value of this parameter can be practically compensated by adjusting the anticipation coefficient. This is examined for κ of 35 veh/mi/ln (20 veh/km/ln) and 65 veh/mi/ln (40 veh/km/ln) using the work zone data. This examination is conducted for relaxation time of 9 seconds, section length of 300 ft , and $\frac{L}{\Delta t}$ of 80 mph. For κ of 35, the calibrated ϑ is 2.77 and for κ of 65 the calibrated ϑ is 3.78. As shown in Figure 4-23, the estimated speeds and queue lengths for these pairs are similar. Average error values for queue length are 56 ft and -80 ft for κ values of 35 and 65, respectively. Average error values for speed are 2.03 mph and 1.82 mph for κ values of 35 and 65, respectively. The maximum queue length happens at the minutes of 37 and 38 for κ values of 35 and 65, respectively. As a result, the two different fine tuning parameters did not show a significant improvement in the queue length data. Thus the study uses κ of 35 because this parameter is mainly used to adjust densities in undersaturated conditions and the amount of adjustment will not be more than the critical density which is equal to 35 pc/mi/ln for our work zone. Furthermore, κ of 35 was used by Cremer and Papageorgiou (1981) for a roadway similar to our work zone. For that roadway, there are no ramps and minute-volume ranges from 500 vphpl to 1750 vphpl; in our work zone there are no ramps and minute-volume ranges from 240 vphpl to 1450 vphpl.



Part a



Part b

Figure 4-23: Estimated a) speed and b) queue lengths for two different fine tuning parameters

4.9.2 Modification 2: Propose two anticipation coefficients

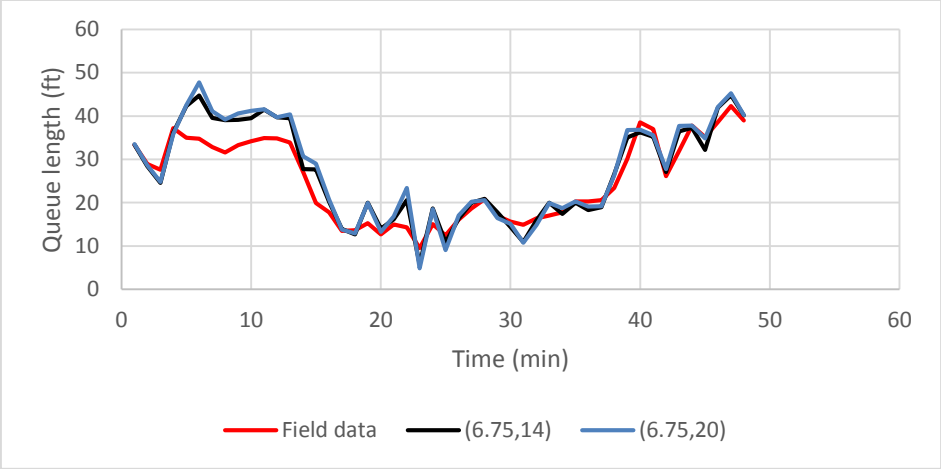
Although the fit based on the Payne model and the modified version of that return admissible average speed error and average queue length error, there is a significant delay in start of queue propagation and there is a sudden queue shrinkage (See Figure 4-23). This suggests using two ϑ values: one for propagation (ϑ_p) and the other one for queue shrinkage (ϑ_s). In the modified model, called Two- ϑ model, the anticipation coefficient is defined as

$$\vartheta_{i,t} = \begin{cases} \vartheta_p & \text{if } D_{i,t} \leq D_{i-1,t} \\ \vartheta_s & \text{if } D_{i,t} > D_{i-1,t} \end{cases} \quad \text{4-12}$$

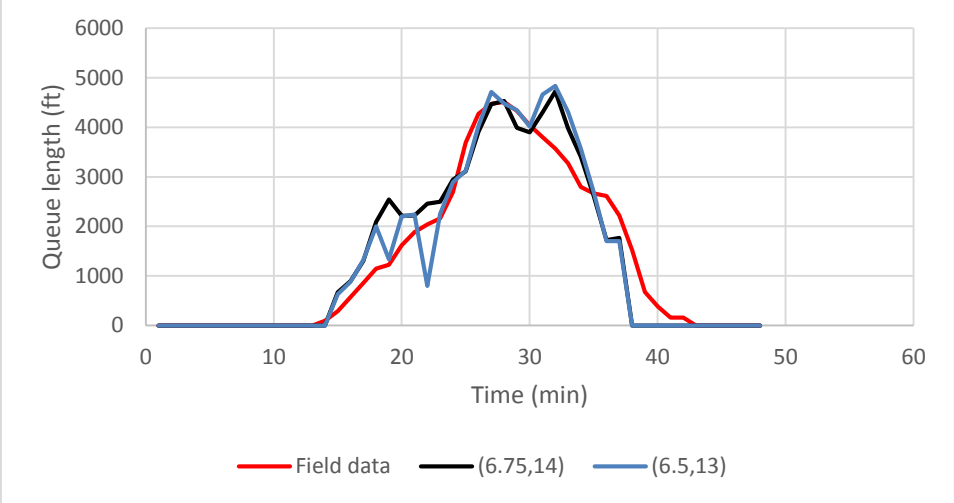
This section explores the effect of two ϑ values when L=450, L/Δt =80 mph, τ =11 sec, and kappa=35. Admissible ϑ values return an absolute average error of less than 2.5 mph and less than 120 ft for speeds and queue lengths, respectively. Initial exploration showed that there could be multiple admissible pairs of ϑs. For instance, the two pairs of (ϑ_p=6.75, ϑ_s =14) and (ϑ_p=6.5, ϑ_s =13) can return good estimates as shown in Figure 4-24. In addition, using 2 ϑ values improved the propagation and shrinkage rates of queue length compared with estimates based on modified Payne model.

Also for these pairs, the ϑ_s is greater than ϑ_p and this can be justified as the denominator of the density gradient term ($\frac{\Delta t * \vartheta}{\tau * L}\left(\frac{D_{i-1,t} - D_{i,t}}{D_{i,t} + \kappa}\right)$) is larger when queue shrinks from the front than when queue propagates. When queue shrinks the denominator $D_{i,t}$ is equal to density in queue and when queue propagates it is equal to density of arriving vehicles which is less than density of queued vehicles. As a result, ϑ_s should be greater than ϑ_p to compensate the effect of larger denominator and to model more realistic queue propagation and shrinkage rate.

A program was written to determine the pairs that potentially can return an average discrepancy of less than 120 ft for queue length and 2.5 mph for speed. The outcomes of this program are the pairs the potentially can return admissible fit and they will be later filtered based on other criteria. The search interval for $\vartheta_p$was from 3 to 8.5 and the search interval for ϑ_s was from 6 to 25. The program found several potential pairs as shown in Figure 4-25 for relaxation times of 7, 9, 11, and 13.



Part a



Part b

Figure 4-24: Estimated a) speeds and b) queue lengths using Two- ϑ model

The range of ϑ_p when τ = 7 sec is from 4.5 to 7.75 and this range moves to the right as τ increases. Similarly, the range of ϑ_s when τ = 7 is from 9 to 15.25 and this range moves upward as τ increases.

To determine the difference between these potential pairs, the acceleration rate (or deceleration rate when it is negative) was computed as below

$$a_{i,t} = \frac{(V_{i,t} - V_{i,t-1})}{\Delta t} * 5280/3600^2$$

This is the acceleration/deceleration rate for a representative vehicle that is in section i, in the beginning of time intervals t and t-1. There is always such a vehicle since Δt and L were

chosen such that vehicles spend more than one time interval in a given section. The $a_{i,t}$ for a typical passenger car should be between -20 ft/sec^2 and 16 ft/sec^2. The number of computations that return an acceleration/deceleration rate outside this range was counted and divided by the total number of computations, 7500. The resulting percentages are shown in Figure 4-26. For each relaxation time, the percentage of over accelerations becomes higher as ϑ_p increases. This is because a highier ϑ_p menases there is highier speed change with respect to a given density change. Thus it is more likely to have more over accelerations for larger parameters. The study selects parameters that return less than 1% of over accelerations. And these parameters are shown in Figure 4-27. The ratio of $\frac{\vartheta_s}{\vartheta_p}$ ranges from 1.86 to 2.63.

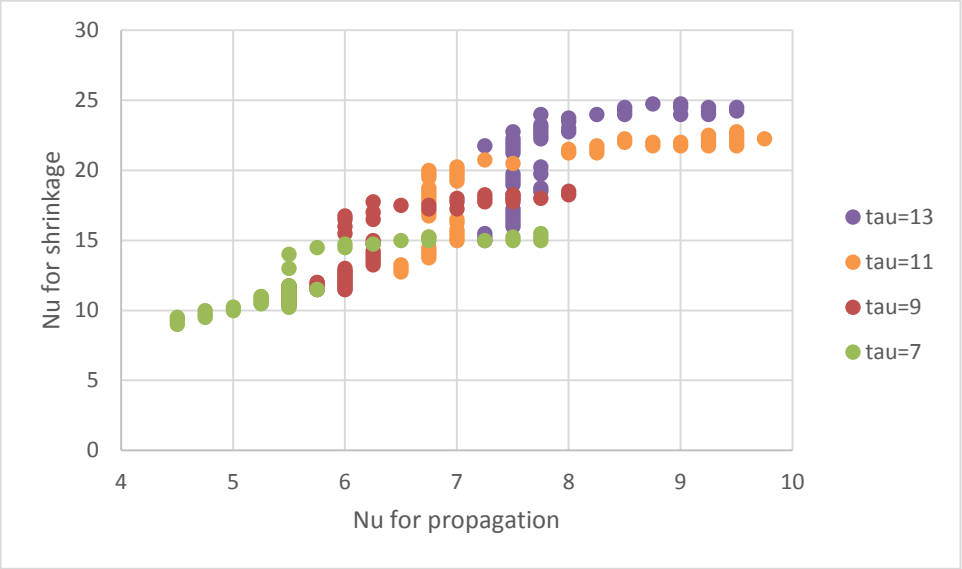


Figure 4-25: Potential admissible pairs of ϑ

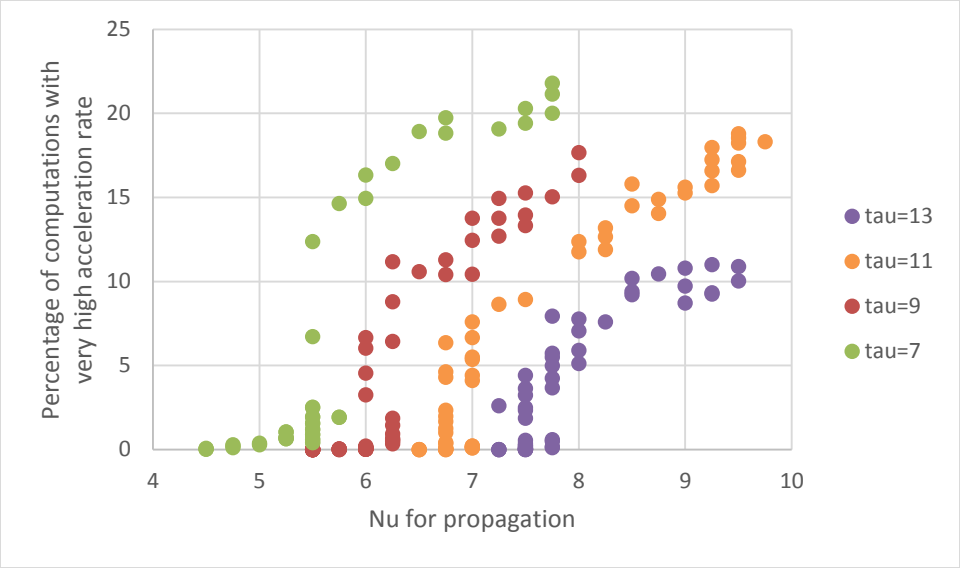


Figure 4-26: Percentage of computations with very high acceleration or deceleration rates

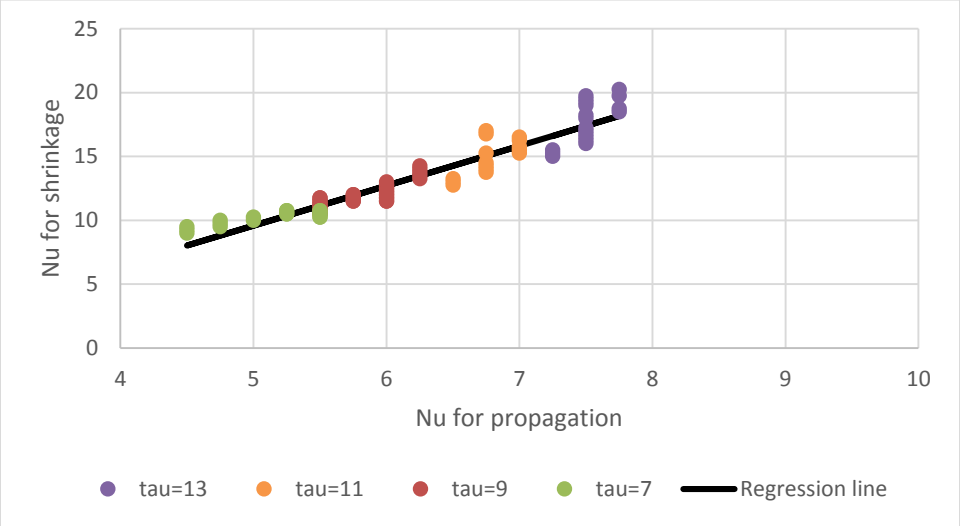


Figure 4-27: Admissible ϑ pairs after filtering based on acceleration and deceleration computations

A regression line is fitted to admissible pairs with the following form:

$$\vartheta_s = a + b\vartheta_p$$

Where, a and b are the regression coefficients.

The domain of ϑ_p that was used in regression is indicted in Table 4-7 for different values of relaxation time.

Table 4-7: Domain of ϑ_p for different values of relaxation time

Relaxation time (sec)	7	9	11	13
Domain of ϑ_p	$4.5 \leq \vartheta_p \leq 5.5$	$5.5 \leq \vartheta_p \leq 6.25$	$6.5 \leq \vartheta_p \leq 7$	$7.25 \leq \vartheta_p \leq 7.75$

Table 4-8 shows the results of regression. The R-square value of 0.905 indicates that a line can sufficiently be used to estimate ϑ_s. The P-values are less than 0.001 implying that both slope and intercept are required to be included in the model.

Table 4-8: Regression results for ϑ_s versus ϑ_p

Number of data points	Parameter values		P-values		R-square
	a	B	a	b	
88	-6.05	3.12	<0.001	<0.001	0.905

4.10 Comparison with first order model

Daganzo (1995a) formulated the finite difference form of the first order model for a general form of a density-flow curve. Based on his study, Equation 2 is reformulated as below for the first order model.

$$V_{i,t+1} = \text{Min}\{R(D_{i-1,t}), S(D_{i,t})\} \qquad \forall i \in I, \forall t \in T \qquad 4\text{-}13$$

Where,

$R(D_{i-1,t})$: The flow rate that can be received by the downstream section.

$S(D_{i,t})$: The flow rate that can be sent out based on the number of the cars present.

The schematic definition of $R(D_{i,t})$ and $S(D_{i,t})$ are shown in Figure 4-28.

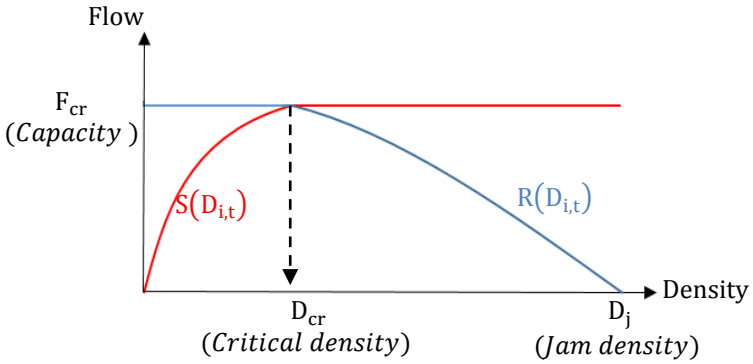


Figure 4-28: Schematic definition of $R(D_{i,t})$ and $S(D_{i,t})$

The first order model does not have any parameters to calibrate. For the same boundary conditions, L=300 ft and $\frac{L}{\Delta t} = 80$ mph, the queue length and speed were estimated. Based on the results shown in Figure 4-29, the queue length was overestimated with average error of 2744 ft. Average error in speed is -1.51 mph with standard deviation of the 7.6 mph.

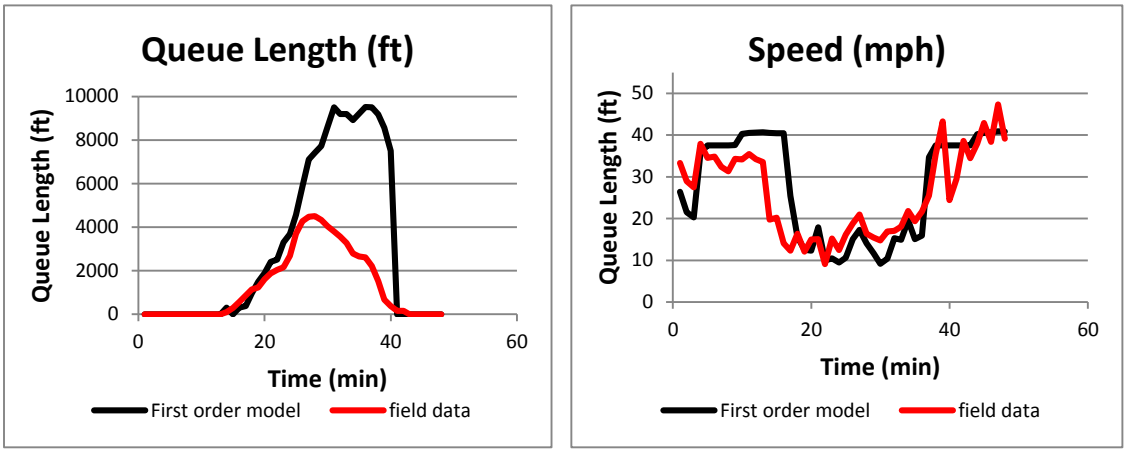


Figure 4-29: Queue length and speed estimated using first order model

The source of discrepancy is investigated by comparison of density, speed, and flow rate estimates versus field data. Figure 4-30 displays density estimates based on the first order model versus field data. The figure shows that during queue conditions in the field (minute 13 to minute 43), the average discrepancy for density is 0.42 pc/mile/ln. The average discrepancy during queue condition estimated by the first order model (minute 16 to minute 40) is 0.17 pc/mile/ln. Thus density computations cannot be the source of the discrepancy in queue conditions.

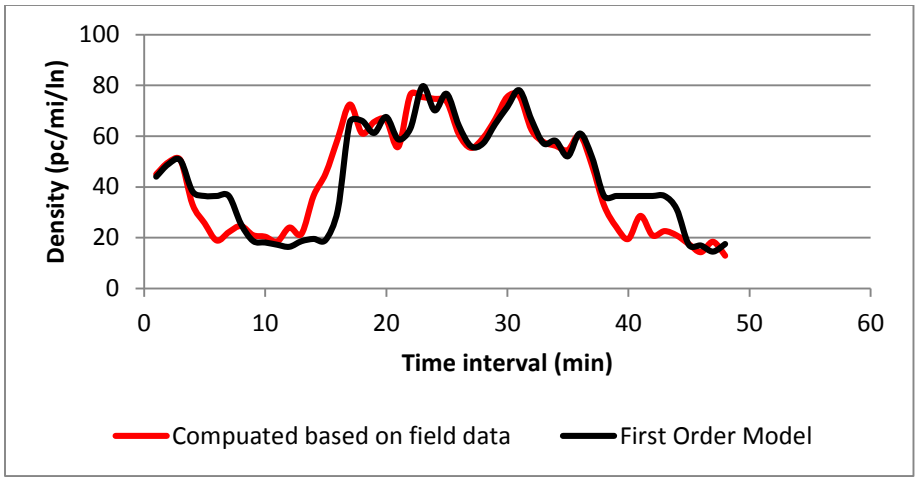


Figure 4-30: Estimated density using first order model versus field data

Figure 4-31 displays speed estimates using first order model versus field data. Average speed when there was a queue in the field (minute 13 to minute 43) is 1.13 mph. The average discrepancy for speed when the first order model estimated a queue length (minute 16 to minute 40) is -1.11 mph. The average discrepancy values are low, but as shown in Figure 4-31Figure 4-32, speed is constantly underestimated from minute 23 to minute 37 with average underestimation of 3.7 mph ranging from 1.4 mph to 6.73 mph. This underestimation is magnified by the corresponding density and causes flow rate underestimation as displayed in Figure 4-31. The average flow rate underestimation for minute 23 to minute 37 is 221 pcphpl. This discrepancy value is close to average speed discrepancy (i.e. 3.7 mph) multiplied by the average density (65.2 pc/ln/) during these minutes.

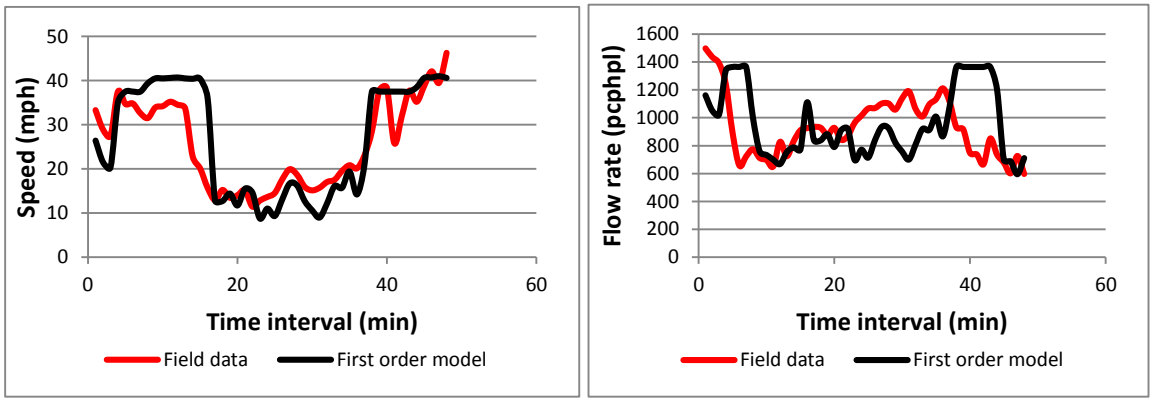
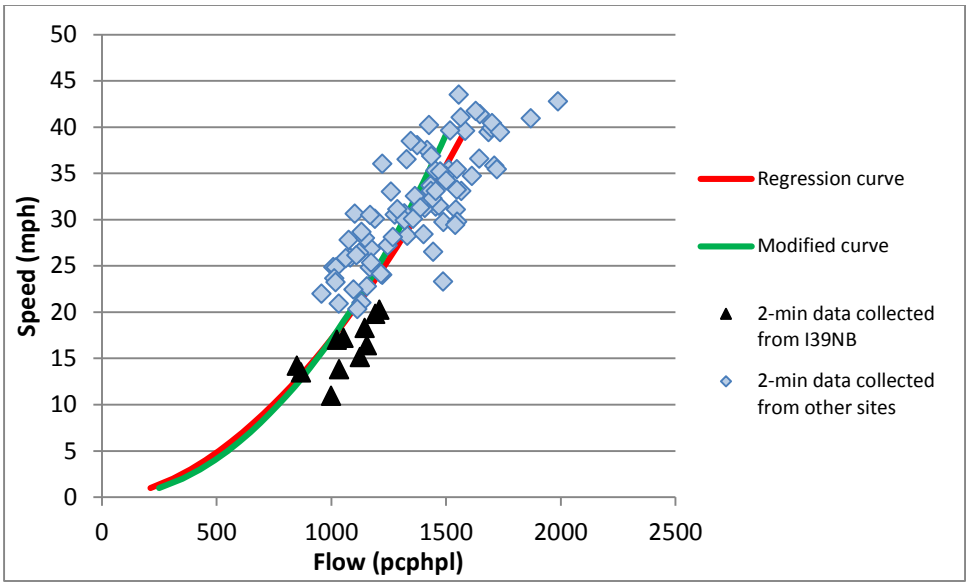
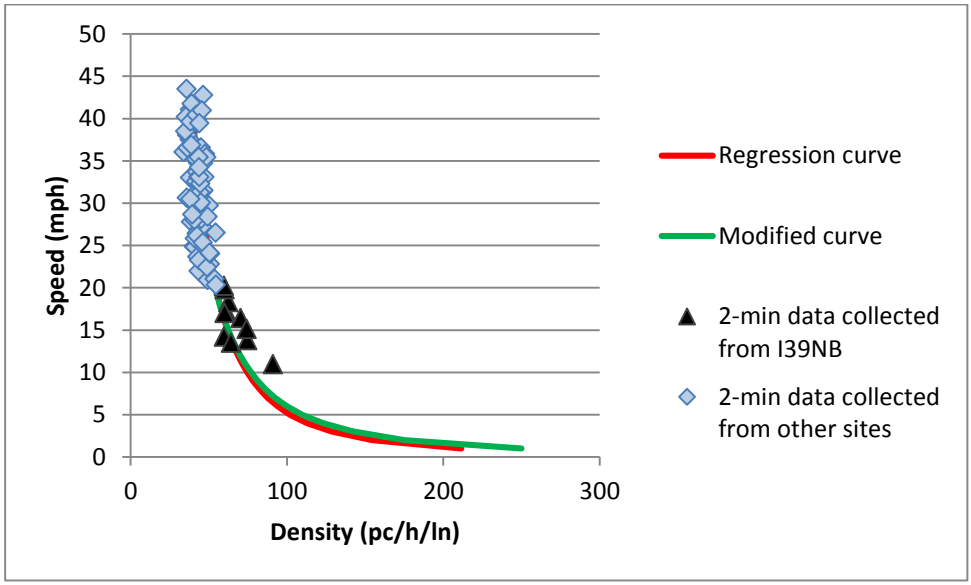


Figure 4-31: Estimated speed and flow rate using first order model versus field data

In order to determine the cause of speed underestimation, field data used to develop speed-flow relationship and corresponding speed-density relationship were plotted in Figure 4-32. Figure 4-32a shows speed and flow data collected from three locations: I39-NB (AM), I80-EB (AM and PM), and I80-WB (AM and PM). A power function was fitted to all the field data repressing a speed-flow relationship for a typical site with flagger and speed limit of 45 mph. The curve slightly modified as explained in Section 5.3 to connect the jam density of 250 pc/mi/ln to the capacity of 1400 pcphpl and speed at capacity of 34 mph determined by field data. The modified curve yields the capacity and the speed at capacity determined from field data and overall is close to regression curve. Figure 4-32b shows the corresponding speed-density data and relationship. Overall speed-flow and speed-density relationships describe whole collected data, but for I39-NB the speed-density relationship returns slower speeds than field data for a given density. One may argue that to evaluate the first order model, the speed-flow relationship or the speed-density relationship should be developed using only I39-NB data because queue length data were collected from I39-NB. In addition, one may argue that similar to second order model that has parameters to calibrate and improve the estimates, parameters should be calibrated for the first order model to improve the estimates. In order to address these concerns, the next section evaluates estimated queue length using the first order model under different cases to improve queue length estimates.



Part a



Part b

Figure 4-32: Congested data and relationships a) speed versus flow b) speed versus density

4.10.1 Modifications of speed-density relationships

4.10.1.1 Case 1) Regression using only I39-NB data

Only I39-NB data collected during queuing conditions are used to develop a regression power function for queuing condition. Nonlinear regression returned the following relationship:

$$\text{Flow} = 297.6 * \text{speed}^{0.4587} \quad \text{4-14}$$

The corresponding speed-density relationship is

$$\text{Speed} = \left(\frac{\text{density}}{297.6}\right)^{\frac{1}{(0.4587-1)}} = \left(\frac{\text{density}}{297.6}\right)^{\frac{1}{(-0.5413)}} \quad (4\text{-}15)$$

Both relationships are shown in Figure 4-33 and Figure 4-34.

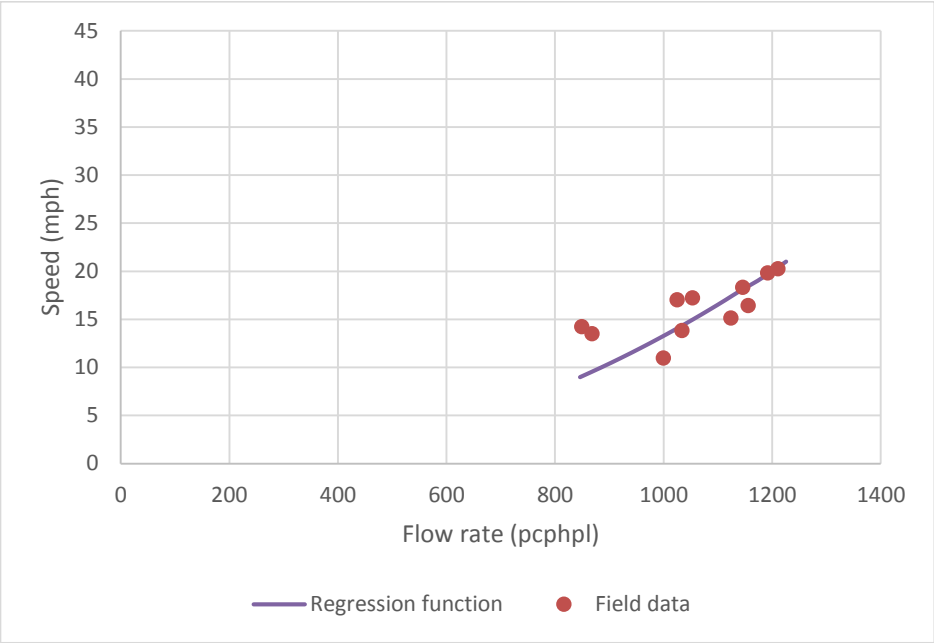


Figure 4-33: Speed versus flow for I39-NB data

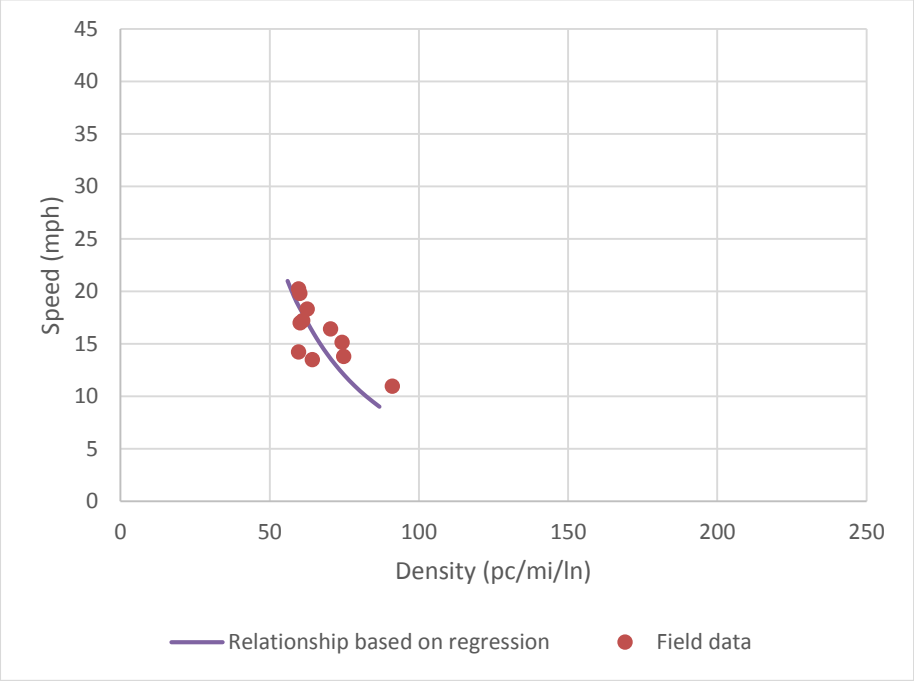


Figure 4-34: Speed versus density for I39-NB data

The regression function does not cover the whole domain of density for congested condition. Extrapolation of the regression function does not return a reasonable jam density or the capacity and the speed at the capacity determined using field data; as a result, a power function is used to connect the regression function to the capacity of 1400 pcphpl and the speed of 34 mph. Another power function is used to connect the regression function to the jam density of 250 and the speed of 1 mph.

Thus speed-density relationship will be in the form of Equation 4-16. The first two sub-functions are the same as those in Equation 4-5 when density is below capacity. The three power sub-functions are used when density is greater than the critical density of 41.18 pc/mi/ln. The regression function was limited to the density values between 59 and 92 which are very close to the minimum (=59.4) and maximum (=91.1) values of density computed based on field data.

$$V(D_{i,t}) = \begin{cases} \dfrac{42.81}{1 + 0.0031 * D_{i,t}} & \text{if } D_{i,t} \le 13.26 \\ \dfrac{62.85}{1 + 0.01863 * D_{i,t}} & \text{if } 13.26 \le D_{i,t} \le D_{cr} = 41.18 \\ \left(\dfrac{D_{i,t}}{437.17}\right)^{\frac{1}{-0.6699}} & \text{if } D_{cr} = 41.18 \le D_{i,t} \le 59 \\ \left(\dfrac{D_{i,t}}{297.6}\right)^{\frac{1}{-0.541}} & \text{if } 59 \le D_{i,t} \le 92 \\ \left(\dfrac{D_{i,t}}{250}\right)^{\frac{1}{-0.461}} & \text{if } 92 \le D_{i,t} \le 250 \end{cases} \qquad \forall i \in I_1, \forall t \in T \qquad \text{4-16}$$

For the updated speed-density relationship, the first order model estimates queue length as shown in Figure 4-35. This shows that back of queue has rapid increase and decrease and this trend is not similar to the queue propagation and dissipation for field data.

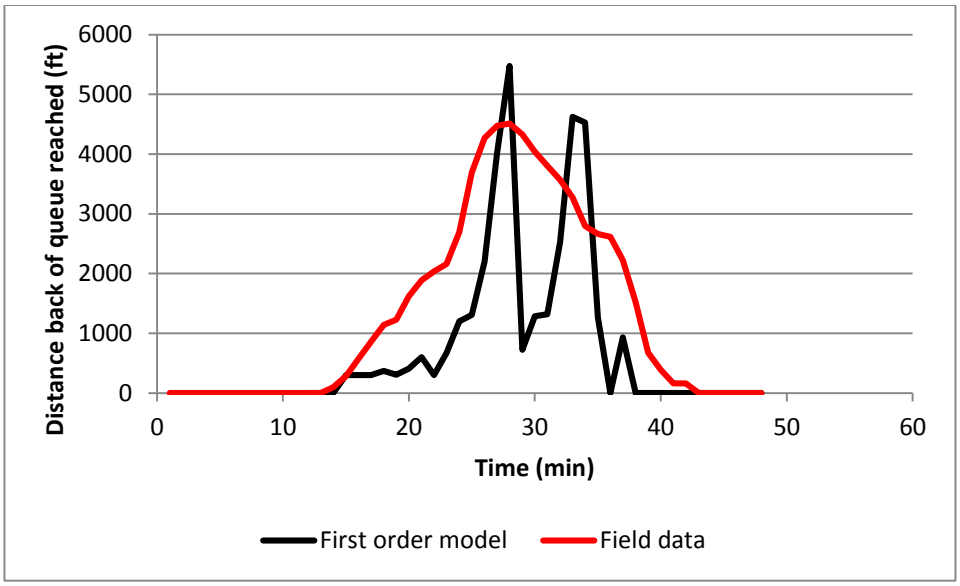


Figure 4-35: Estimated queue length using first order model for Case 1

To investigate the reason for this trend, the study explores average speed computed for each minute and each section shown in Figure 4-36. Each column represents a minute during queuing condition which starts from minute 16 and ends at minute 35. Each row represents a section. The Row 1 is the most downstream section. If average speed drops below 20 mph, it is assumed that the section is in queuing condition. The threshold of 20 mph is the same as the threshold used for the second order model. The cells in queuing conditions are highlighted in red. One of the reasons to have fluctuation in back of queue could be due to threshold of 20 used to declare queuing condition. For example in minute 29, there are two sections (the Sections 1 and 2) under queuing condition while speeds of the Sections 3 to 18 are below 30 mph and one may consider these sections under queuing condition because corresponding speeds are below the critical speed. Thus in the next section, the threshold to declare queuing condition will be modified.

Section numbers	Time (min)																			
	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
1	17.8	13.5	13.7	15.2	12.7	15	14.3	9.48	15	12.4	16.1	18.6	20.8	17	15.7	14.9	16.3	17	17.7	20.3
2	39.3	37.8	26.8	34.1	22.2	17.3	36.1	16.7	14.9	12	14.2	19.4	21.9	19.1	15.8	12.7	12.8	18.2	20.3	21.1
3	39.1	37.8	36.6	36.4	35.8	35	38	27.1	13.6	13.4	13	18	21.3	20.3	16.5	13.7	12.1	16.5	20.5	19.9
4	38.9	37.6	36.6	36.4	36.2	37.1	37.8	35.1	17.6	14.3	12.4	16.5	20.6	21.1	17.2	14.6	12.1	14.9	20	19.6
5	45.8	44.5	43.7	43.5	43.4	44.5	44.6	42.8	41.7	22.4	18.3	22.5	27.9	28.8	24.8	21.5	18.2	20.5	27.2	27.2
6	45.7	44.4	43.7	43.4	43.4	44.6	44.4	42.9	43.5	35.4	18.4	21.5	27.6	28.9	25.3	22	18.6	19.7	26.6	27.3
7	45.6	44.2	43.7	43.4	43.5	44.8	44.1	42.9	43.5	41.8	19.3	20.6	27.1	29	25.8	22.4	19.1	19	25.9	27.4
8	45.5	44.1	43.7	43.3	43.5	45	43.8	43	43.5	42.9	23.3	19.8	26.5	28.9	26.3	22.8	19.6	18.6	24.9	27.5
9	45.4	43.9	43.7	43.3	43.6	45.2	43.5	43.2	43.5	42.8	27.9	19.2	25.7	28.8	26.8	23.2	20.2	18.4	23.9	27.6
10	45.3	43.9	43.7	43.3	43.6	45.3	43.2	43.3	43.4	42.6	31.7	18.9	24.8	28.7	27.3	23.6	20.6	18.3	22.7	27.7
11	45.2	43.8	43.6	43.3	43.7	45.4	43	43.4	43.4	42.5	34.8	18.9	23.8	28.5	27.7	24	21.1	18.5	21.6	28.8
12	45	43.7	43.6	43.3	43.8	45.4	42.8	43.5	43.4	42.4	37.4	19.2	22.7	28.3	28.1	24.4	21.5	18.7	20.6	32.1
13	44.9	43.7	43.6	43.3	44	45.2	42.7	43.5	43.3	42.2	39.4	20	21.7	28	28.4	24.8	21.9	19.1	19.7	35.3
14	44.8	43.7	43.5	43.3	44.1	45.1	42.7	43.6	43.3	42.1	41	21.4	20.7	27.6	28.6	25.2	22.3	19.5	19.1	38.9
15	44.6	43.7	43.5	43.3	44.3	44.8	42.7	43.6	43.2	42	41.3	24.1	20	27	28.8	25.6	22.6	19.9	18.9	44.8
16	44.5	43.7	43.4	43.4	44.5	44.5	42.8	43.6	43.1	41.9	41.3	27	19.6	26.2	28.8	26.1	23	20.3	31.6	47.1
17	44.3	43.7	43.4	43.4	44.7	44.2	42.9	43.5	43	41.8	41.3	29.7	19.4	25.3	35.2	29.5	23.4	23.3	45.7	47.1
18	44.1	43.7	43.4	43.5	44.9	43.9	43	43.5	42.9	41.8	41.3	32.3	19.6	26.3	42.8	33.5	23.7	31.7	46.4	47.1
19	44	43.7	43.3	43.5	45.1	43.6	43.2	43.5	42.7	41.7	41.3	40.3	26	38.7	43.2	36.9	24.1	38.7	46.5	47
20	43.8	43.7	43.3	43.6	45.2	43.3	43.3	43.4	42.6	41.6	41.3	42.4	44.5	44.5	43.1	39.4	24.5	44.7	46.6	46.8
21	43.8	43.7	43.2	43.6	45.4	43	43.4	43.4	42.5	41.5	41.3	42.7	44.6	44.3	42.9	41	28	47.3	46.6	46.6
22	43.7	43.6	43.2	43.7	45.6	42.7	43.5	43.4	42.3	41.4	41.3	42.9	44.7	44.2	42.7	41.5	34.5	47.2	46.7	46.4
23	43.7	43.6	43.2	43.8	45.5	42.5	43.6	43.4	42.2	41.4	41.3	43.2	44.8	44	42.5	41.5	41	47.1	46.8	46.2
24	43.7	43.6	43.3	44	45.3	42.5	43.6	43.3	42.1	41.3	41.3	43.5	44.9	43.9	42.3	41.4	45.3	46.9	46.8	46

Figure 4-36: Spatiotemporal distribution of estimated average speed the Case 1

4.10.1.2 Case 2) Modifying queuing thresholds

Based on the discussion above, the speed threshold to declare queuing condition is modified. A given section is in queuing condition if the corresponding speed is below the critical speed. The Sections 1 to 4 have a speed-density relationships corresponding to flagger presence and its critical speed is 34 mph. The other sections have a speed-density relationship corresponding to no flagger presence with the critical speed of 41 mph. Thus instead of using a single threshold of 20 mph for all sections, two thresholds are used: 34 mph for the Sections 1 to 4 and 41 mph for the Sections 5 to 24. The spatiotemporal distribution of estimated speed is shown in Figure 4-37 for the new thresholds. The cells under queuing condition are highlighted in red for the Sections 1 to 4 and in orange for the Sections 5 to 24. This figure shows that the sections between back of queue and front of queue are under queuing conditions, thus the distance back of queue reached represents the queue length. The corresponding queue length is displayed in Figure 4-38 showing a more realistic queue propagation and dissipation pattern than Figure 4-35 for the Case 1. The average estimated queue length for the Case 2 is 2178 ft which is 123 ft longer than 2055 ft which is the average queue length for the field data. The maximum queue length for the Case 2 was overestimated by 1787 ft compared with field data. Although the estimated average queue length is close to the field data, the maximum queue length is considerably longer than field data. The next sections will modify values of some parameters to improve the estimated maximum queue length.

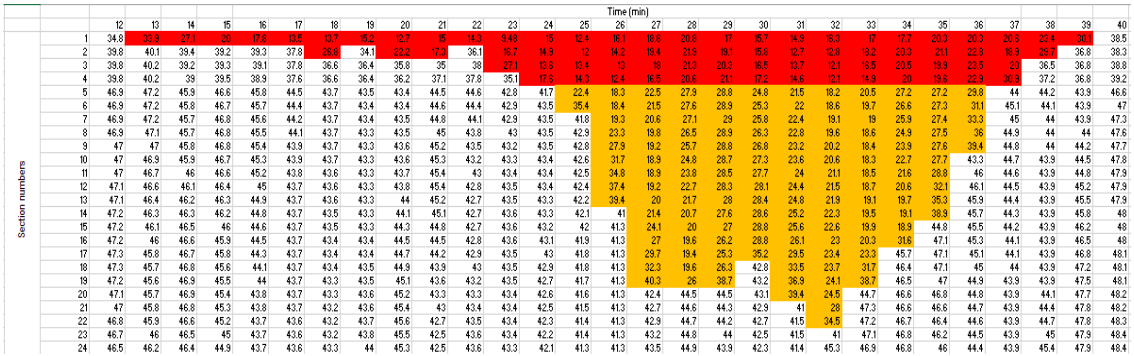


Figure 4-37: Spatiotemporal schematic of estimated average speed using first order model for Case 2

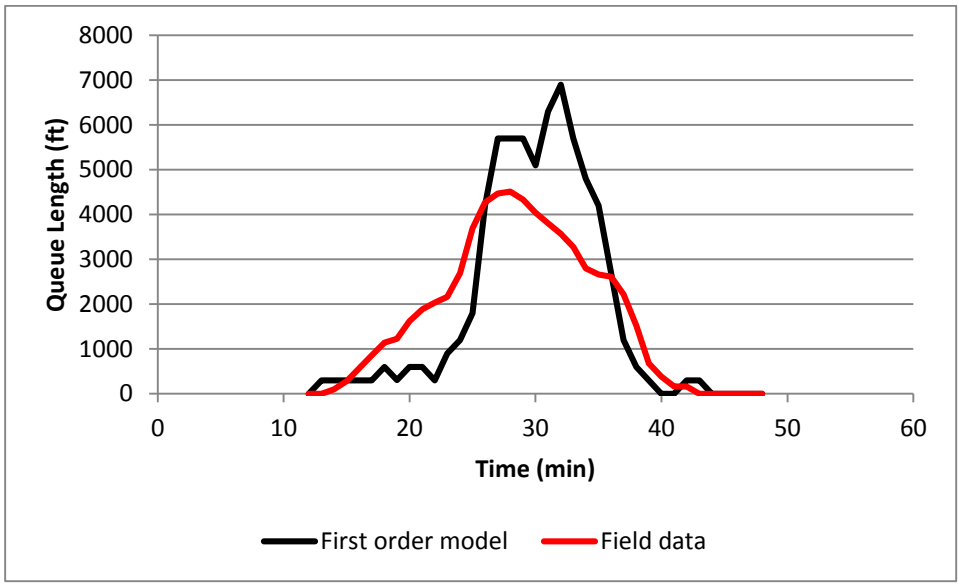


Figure 4-38: Estimated queue length using first order model for Case 2

4.10.1.3 Case 3) Modifying free flow speed

There are two free flow speed values that potentially could be modified. 1) Free flow speed for flagger presence (The Sections 1 to 4) and 2) Free flow speed for no flagger presence (The Sections 5 to 24). The study does not change free flow speed of 42.8 mph for flagger presence because it was estimated based on field data. For no flagger presence, there was no field data to determine the free flow speed and the study has been using a free flow speed based on prevailing conditions. The study will modify this free flow speed to improve the queue length estimates. The posted speed limit for the Sections 5 to 24 was 45 mph. The free flow speed is usually several miles per hour greater than the posted speed limit. So far, the free flow speed of 50 mph has been used in the speed-density relationship for Sections 5 to 24. In this section, free

flow speed is increased from 50 mph to 52 mph. Increasing free flow speed from 50 mph to 52 mph returned average queue length of 1878 ft which is 177 ft shorter than the average queue length of 2055 ft for the field data. The maximum estimated queue length is 5700 ft which is 1187 ft longer than the maximum queue length in the field. The estimated maximum queue length is closer to field data than the estimated value based on free flow speed of 50 mph in Case 2. This means increasing free flow speed from 50 mph to 52 mph improved estimated maximum queue length; however the average queue length is underestimated by 177 ft. Thus, further increase of free flow speed may reduce maximum queue length, but it can considerably reduce average queue length which is not desirable. In addition further increase of free flow speed may not describe the prevailing conditions because speed monitoring display was present in the field and some drivers reduced speed in response to speed monitoring display; as a result it is not likely to have average speed of drivers to be greater than posted speed limit plus 7 mph.

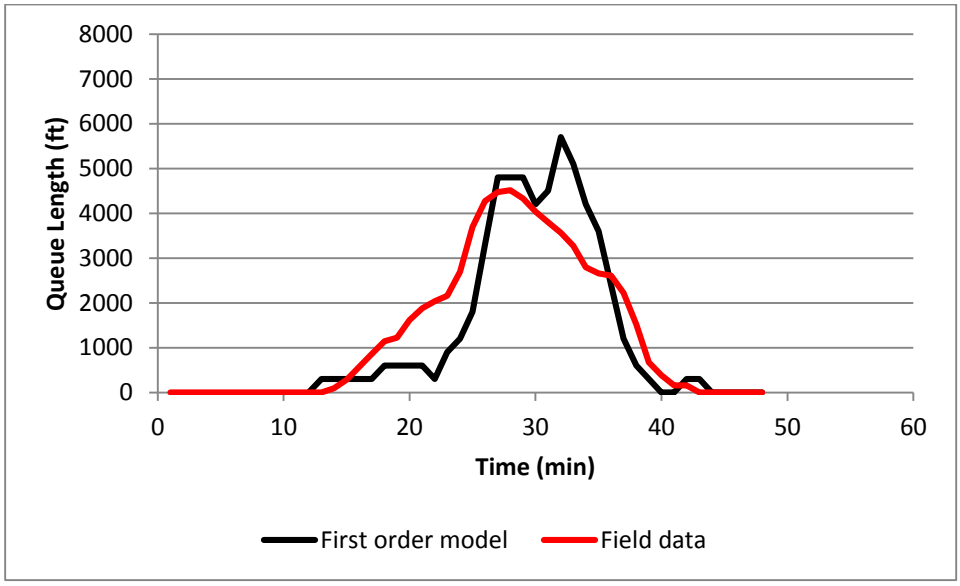


Figure 4-39: Estimated queue length using first order model for Case 3

4.10.1.4 Case 4) Modifying jam density

To determine effect of jam density on the estimated queue length, the jam density was changed from 250 pc/mi/ln to 200 pc/mi/ln. The queue length corresponding to jam density of 200 pc/mi/ln is shown in Figure 4-40. The effect of changing jam density on queue length was relatively minor compared with the Case 3. The estimated average queue length is 1868 ft which is only 10 ft shorter than the estimated average queue length for the Case 3. The estimated

maximum queue length is 5700 ft which is equal to that in the Case 3. Thus in this case reducing jam density to 200 pc/mi/ln did not significantly influence the results. Jam density values below 200 pc/mi/ln is not evaluated because 200 pc/mi/ln is corresponding to spacing of 26.4 ft. between passenger cars including 19 ft for a car length and 7.4 ft for buffer length in stopped queue, and it is not likely to have an average buffer length of greater than 7.4 ft in a stopped queue. On the other hand, jam density values greater than 250 pc/mi/ln is not evaluated because they are corresponding to average spacing of shorter than 21 ft and average buffer length of shorter than 2 ft. These short average buffer lengths between passenger cars are not likely to observe.

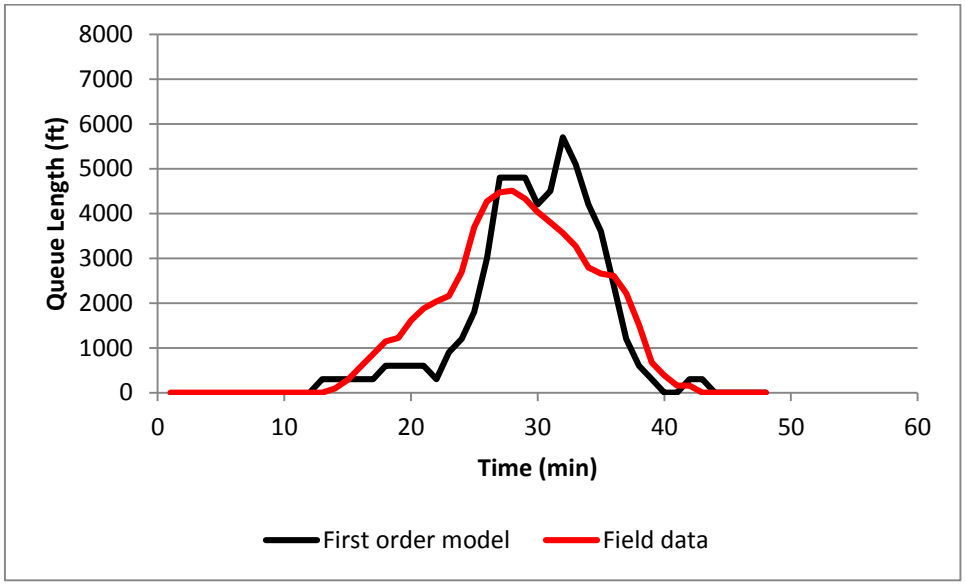


Figure 4-40: Estimated queue length using first order model for Case 4

4.10.1.5 Conclusions

First order model in the Case 3 returned better results than other cases. In particular it underestimated average queue length by 177 ft which is corresponding to 8.6% error, and it overestimated maximum queue length by 1187 ft which is corresponding to 26.3% error. The standard deviation of error in queue length estimates was 955 ft which is corresponding to coefficient of variation of 46.5%.

The second order model returned better average and maximum queue lengths than the first order model, but a larger coefficient of variation. The discrepancy between the second order model estimates and the field data depended on the values of τ and ϑ. For example for τ =11 and

ϑ =1.80, the second order model underestimated average queue length by 100.5 ft which is corresponding to 4.9% error. The maximum queue length was overestimated by 309 ft which is corresponding to 6.8% error. The standard deviation of error was 1734.7 ft which is corresponding to coefficient of variation of 84.4%.

The Two- ϑ model retuned better average queue length, maximum queue length, and coefficients of variation than the first order model and the second order model. The amount of discrepancy between estimates and field data depended on parameter values of τ, ϑ_p, and ϑ_s. For example for τ=11, ϑ_p=6.5, ϑ_s=13, average error in the queue length was 32 ft which is corresponding to 1.56% error. The maximum queue length was overestimated by 318 ft which is corresponding to 7.1% error. The standard deviation of error is 614.3 ft which is corresponding to coefficient of variation of 29.9%.

Thus based on the numerical comparisons, it can be concluded that the Two- ϑ model is better than the first order model and the second order model. One may use the Two- ϑ model when the rate of queue shrinkage and dissipation are significantly different. Consequently, one may use the second order model when the rates of dissipation and shrinkage are similar. In addition, in optimization programs using the second order model can significantly reduce computational complexity compared with the Two- ϑ model which practically has a discontinuous function for ϑ. Thus this study uses the second order model to reduce computational complexities in the optimization program. The next section will use findings of this chapter to determine the second order parameters for the optimization program.

4.11 Finding parameters of the second order model for the optimization program

The second order model was calibrated using field data for section lengths up to 450 ft and time increments up to 5.11 seconds. These section lengths and time increments are short and result in large number of constraints and variables requiring huge computational resources to solve the optimization problems. To reduce the computational complexities, the study will use a longer section length of 0.25 miles. This section length is still shorter than the section lengths of 0.31 miles (0.5 km) and 0.625 miles (1 km) used by Carlson et al. (2010) and Hegyi et al. (2005). The section lengths of 0.25 miles would result in more accuracy in determining location of the critical CMSs than longer section lengths. The only remaining issue is that field data is not sufficient to calibrate the second order model for section length of 0.25 miles because extend of

congestion was limited (less than a mile). To overcome this issue, the study creates simulated data using the Two-ϑ model and uses the simulated data as a base to calibrate the second order model. More details of the methodology is presented in the next section. .

4.11.1 Methodology

In this section, the roadway is similar to that introduced in section 4.2, but the section with speed limit of 45 mph and no flagger is extended to have the total length of 2 miles. The reason for extension is that there should be enough space for a queue length of 1 mile or longer and this will allow evaluating several sections under queuing conditions.

The methodology has two parts:

Part 1) For a given boundary and initial condition, the queue length is found using section length of 450 ft for which the Two-ϑ model has been calibrated.

Part 2) For the same boundary and initial condition, the second order model is calibrated for the section length of 0.25 miles to compute similar queue length estimated in part 1. It should be noted that for the section length of 0.25 miles, the second order model has only one ϑ and the Two-ϑ model is not used because incorporating the Two-ϑ model in the optimization requires integer variables, and this increases problem complexity.

4.11.1.1 Part 1) Queue length for the short section length

For computational purposes, the study selects the section length of 0.1 miles which is slightly larger than 450 ft for which the Two-ϑ model has been calibrated. Thus the parameters can be selected based on the guideline developed for section lengths of 450 ft. The logic to select the parameters is as follow:

- Based on the guideline to select relaxation time in Section 4.8, the admissible range for relaxation time is between 10 seconds and 13 seconds. The selected value is 11 seconds since it is a middle value in the range and the study has already calibrated the Two-ϑ model for this value of the relaxation time.
- Based on the results of calibration for the Two-ϑ model, I select 6.5 $\frac{\text{mile}^2}{\text{hr}}$ for ϑ_p , and 12.75 $\frac{\text{mile}^2}{\text{hr}}$ for and ϑ_s. These values are among the calibrated values for reaction time of 11 seconds, and they are less likely to overestimate acceleration or deceleration rates.

- The value of Δt was selected to be 4.5 seconds because it returns $\frac{L}{\Delta t}$ of 80 mph for which the Two-ϑ model is calibrated.
- Kappa is selected to be 35 pc/mi/ln which is the same value used in the Two-ϑ model.

Using these parameters, the queue length is estimated when arriving volume is 1200 pc/hr, and after 1 min from the beginning of the analysis, capacity drops to 900 pc/hr for 15 minutes. This capacity drop resulted in maximum queue length of 5573 ft and average queue length of 3594 ft, in the simulated data. The next section will find the parameters that estimates this queue pattern.

4.11.1.2 Part 2) Calibration for the long section length

The second order model is calibrated when only one ϑ is used. The Two-ϑ model is not used because it will require integer variables in the optimization problem causing significant computational effort. The study selects the parameters using the following logic:

- Based on the guideline in 4.8, relaxation time can be selected proportional to the section length; thus the relaxation time of 11 is multiplied by 2.5 (the ratio between the section lengths of 0.25 miles and 0.1 miles) to find the relaxation time. The resulting value will be 27.5 seconds, and it is rounded to 28 seconds.
- The Δt of 10 seconds is selected since it is a round number and is appropriate for practical purposes. In addition, it returns $\frac{L}{\Delta t}$ of 90 mph which is close to 80 mph, the corresponding ratio for section length of 0.1 mile.
- The same kappa value of 35 is used.

The ϑ value is selected such that it returns the closest queue length trend to that estimated for the section length of 0.1 miles. The Figure 4-41 shows how queue length changes for different values of ϑ. The ϑ values of 17 and 18 underestimate average queue length by 285 ft and 93 ft, respectively. The ϑ values of 19 and 20 overestimate average queue length by 130 ft and 429 ft, respectively. The ϑ values of 18 and 19 return lower error values for average queue length. Between these two values, the study uses ϑ value of 18 because it creates less discrepancy in maximum queue length and congestion duration.

Thus, the values of the second order model parameters used for this study are reported in Table 4-9.

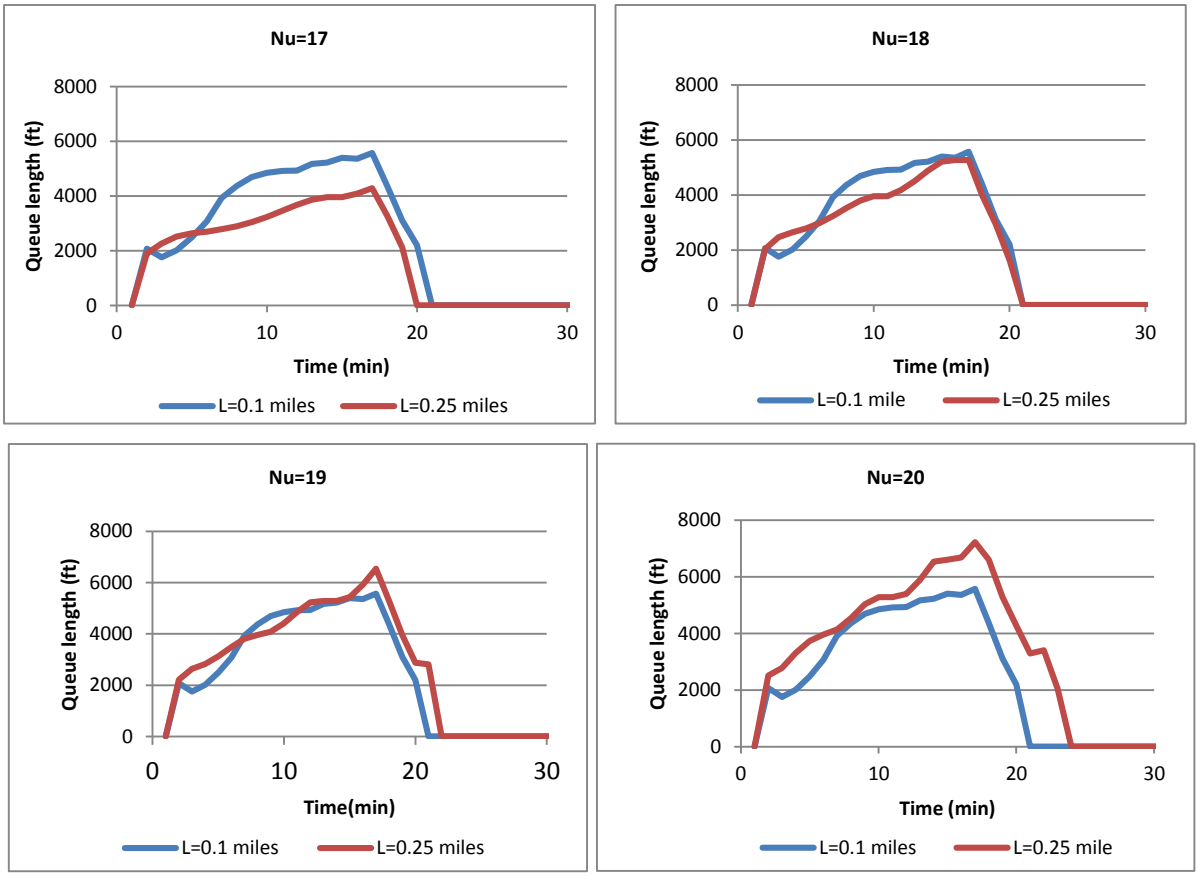


Figure 4-41: queue length estimated for sections lengths of 0.1 miles and 0.25 miles using different ϑ (nu) values

Table 4-9: Second order parameter values used in the optimization program

Parameter	L (miles)	Δt (sec.)	τ (sec.)	$\vartheta(\frac{Mile^2}{hr})$	κ (veh/minle/ln)
Value	0.25	10	28	18	35

Chapter 5 A continuous transformation method to optimize speed harmonization with piecewise speed-density relationship

The past optimization studies on speed harmonization (Kang et al 2004, Berton et al 2002, Kotsialos et al 1999) mostly used one-regime speed-density relationships such as Greenshields-type models or exponential models, which would reduce problem complexity and help solving optimization programs. However as it will be discussed later, one regime model may not be sufficient to describe traffic flow at oversaturated conditions. As a result, the expected performance improvements due to speed harmonization may not be observed if a single-regime model is used. In addition, recent traffic stream models, (e.g. the relationships proposed in Highway Capacity Manual 2010) are not supporting a single-regime speed-density model, rather they are piecewise and nonlinear. However, inclusion of a piecewise and nonlinear model in an optimization program causes more computational complexities. In particular, incorporation of piecewise models in optimization programs usually requires integer variables that would cause high level of complexity. The computational issue becomes more pronounced when integer variables are combined with nonlinear constraints, which could be the case for speed harmonization with nonlinear speed-density models. In this case, the problem becomes a Mixed Integer Nonlinear Program (MINLP), which is difficult to solve especially for large-scale problems. As a result, a methodology is needed to reduce computational difficulties when piecewise nonlinear speed-density relationship is incorporated in the optimization program.

Thus, the objective of this chapter is to develop and solve a mathematical program, called WZ-Speed Harmonizer (WZSH), to optimize speed harmonization with piecewise speed-density relationships. The study shows that the WZSH can be equivalently formulated as a Nonlinear Program (NLP) when piecewise speed-density functions have certain mathematical properties. These properties may not exist in the traffic stream models that are developed so far, but as shown later, one can approximate these models with some functions which satisfy the desired properties. Formulation of the approximate functions to eliminate the integer variables and to considerably reduce the problem complexity is discussed in the following sections.

5.1 Motivation: Issues with one-regime traffic stream models

One-regime traffic stream models are simple and easy to be formulated in optimization programs; however they may not be sufficient to model congested conditions. Alternatively,

multi-regime models (i.e. piecewise models) have been developed to more accurately describe congested traffic conditions. Early examples of multi-regime relationships are Edie's (1961) and Dick's (1966) speed-density models that are composed of two equations for uncongested and congested conditions. HCM also proposed different versions of piecewise speed-flow relationship from 1985 to 2010. These curves are widely used in practice to analyze level of service in US freeways and highways.

To highlight the insufficiency of one-regime models for congested conditions, the next section compares an HCM 2010 speed-flow relationship with single-regime models, which have been used in optimization programs for speed harmonization, as discussed in the next section. This comparison will show how single regime models can significantly cause discrepancy for congested condition although they can return acceptable estimates for uncongested conditions. This sets up the motivation to develop an efficient and fast methodology to solve optimization program with multi-regime models in this chapter.

5.1.1 Approximation of multi-regime models using one-regime models

Exponential equations, defined in Equation 5-1, have been used as a single regime fundamental diagram in speed harmonization studies (Berton et al 2002, Carlson et al 2010).

$$V(D) = V_f \exp\left(\frac{-1}{a}\left(\frac{D}{D_{cr}}\right)^a\right) \quad 5\text{-}1$$

Where,

D: Road density (pc/mi/ln),

V(D): Speed (mph) corresponding to density D,

V_f: Free flow speed (mph),

D_{cr}: Critical density (i.e. density at capacity), and

a: Parameter of the equation.

This section will show that using Exponential equation for a freeway may cause significant inaccuracy in the congested conditions. The parameters of the exponential models are computed such that they return the same free flow speed, capacity and critical density as an HCM 2010 speed-flow relationship returns. As an example, we choose HCM 2010 speed-flow relationship when free flow speed is 65 mph. The corresponding capacity and critical density are

2350 pcphpl and 45 pc/mi/ln, respectively. Based on the definition, V_f is equal to 65 mph which is the free flow speed; D_{cr} is equal to 45 pc/mile/ln which is the critical density; the parameter a is computed such that the Exponential model returns a capacity of 2350 pcphpl. Capacity is computed as:

$$\text{Capacity} = D_{cr} * V(D_{cr})$$

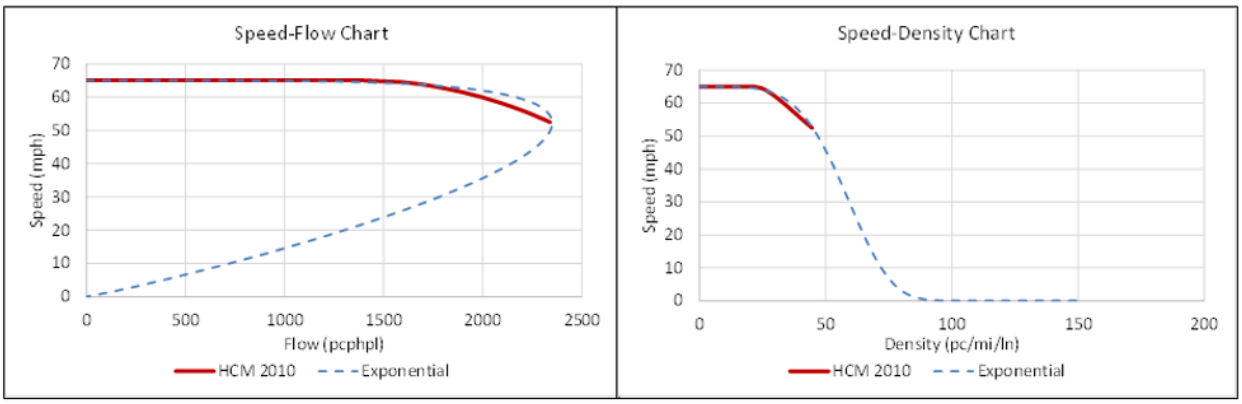
Based on Equation 5-1, $V(D_{cr})$ will be equal to $V_f \exp(\frac{-1}{a})$. As a result:

$$\text{Capacity} = D_{cr} * V_f \exp(\frac{-1}{a})$$

$$a = \frac{-1}{\ln(\frac{\text{Capacity}}{D_{cr} * V_f})}$$

Based on the equation above, the value of parameter "a" will be 4.569.

Figure 5-1a shows the speed-flow curves for the HCM 2010 model and the Exponential model. It demonstrates a good match when traffic is in undersaturated conditions. The HCM 2010 relationship does not have oversaturated part and overall behavior of the Exponential approximation does not reveal counterintuitive trend; however, the speed-density curves, in Figure 5-1b, show that the Exponential approximation results in a low estimate of (practical) jam density. If jam density is practically defined as density when speed is 1 mph, then the exponential approximation roughly returns jam density of 90 pc/mi/ln, which is corresponding to the spacing of 58 ft. This is considerably higher than 25 ft which could be considered as the spacing between passenger cars in stopped queue. The very low density for speed of 1 mph is an indication that the rate of speed reduction is very high in oversaturated condition. Thus exponential equation 5-1 may not be a good choice for some of freeways.



Part a

Part b

Figure 5-1: HCM 2010 fundamental relationships approximated by exponential Equation a) speed-flow relationships b) speed-density relationships

Another model that is used in speed harmonization studies is Greenshields model in which jam density is two times as big as the critical density in the curve. For HCM 2010 speed-flow relationships the critical density is 45 pc/mi/ln, thus the jam density is 90 pc/mi/ln which is a very low estimate. Similar the exponential curve this jam density is corresponding to spacing of 58 ft that is large for sopped queues.

As a result, one regime models may not be sufficient to describe all traffic conditions. For those cases that one-regime models are not sufficient, piecewise speed-density models should be included in traffic control problems; however piecewise models usually requires integer variables to be included in the optimization program, and this causes longer solution time and requires more computational resources; thus a methodology needs to be developed to solve the optimization programs with low computational costs.

5.2 Formulation of piecewise speed-density functions

In the formulation presented in Chapter3, there is no assumption whether the speed-density function is one-regime or multi-regime (piecewise). In this section the formulation is developed for the piecewise model.

5.2.1 MINLP formulation

Assume that the general form of a piecewise continues speed-density function is expressed with a finite number, K, of sub-functions as:

$$V(D) = \sum_{k=1}^{K} R_k(D)\ I_{[D_{k-1},D_k]}(D) \qquad 5\text{-}2$$

Where,

$R_k(D)$'s are twice differentiable, and continuous functions of density and can be nonlinear. For the sake of simplicity, time and section indexes for density (D) are not shown.

D_{k-1} and D_k determine the density boundaries for the kth regime. In particular, $D_0 = 0$, and $D_K = D_{jam}$(i.e. jam density) .

$I_{[D_{k-1},D_k]}(D)$s are indicator functions that return 1 when $D \epsilon [D_{k-1}, D_k]$; otherwise it is zero.

A piecewise function can be modelled as a set of disjunctive constraints. A common approach to formulate disjunctive constraints is Big-M method which could be implemented as

$$D_{k-1} - (1 - x_k)M \leq D < D_k + (1 - x_k)M \qquad \forall k \in \{1, .., K\} \qquad 5\text{-}3$$

$$R_k(D) - (1 - x_k)M \leq V(D) < R_k(D) + (1 - x_k)M \qquad \forall k \in \{1, .., K\} \qquad 5\text{-}4$$

$$\sum_{k=1}^{K} x_k = 1 \qquad 5\text{-}5$$

Where,

x_k: Binary variables determined as

$$x_k = \begin{cases} 1 & \text{if } D_{k-1} \leq D < D_k \\ 0 & \text{otherwise} \end{cases}$$

M: is a sufficiently big number. It could take different values for different constraints, but for simplicity it is assumed to be a fixed number.

The inequality 5-3 determines the traffic regime based on the value of density and constraints 5-4 find the corresponding value of speed. The formulation above can be more simplified as one may combine constraint 5-5 with other constraints and eliminate one integer variable, say x_K. As a result, the piecewise function formulation includes K-1 integer variables for a single section and a single time interval using. Assuming that the analysis period is divided into N_t time intervals and the roadway is divided into N_s sections, the number of integer variables would be (K-1)* N_t*N_s. As will be discussed later, this number could be a large number for practical problems and solving the corresponding problems would be typically time

consuming and difficult. Thus, an alternative approach which results in less computational efforts is favorable especially for large scale problems or real time applications.

5.2.2 From MINLP to NLP

Two alternative approaches could be considered to simplify the problem solution.

1) One approach is to approximate the piecewise nonlinear function with a piecewise linear function. Generally, this approach may be useful when it transforms the MINLP into a Mixed Integer Linear Program (MILP); however in this problem, it does not eliminate nonlinearity that exists in Constraints 2 and 4; thus the formulated speed harmonization problem is still MINLP. Moreover, it is likely that number of integer variables in MILP is higher than that in MINLP. In particular, one may use more linear sub-functions to capture the nonlinearity in the original sub-functions, and this requires more integer variables which may not be desirable to be included in the program.

2) Another approach is to transform the MINLP to an NLP. The transformation eliminates the integer variables, and continues methods can be applied to solve the problem. This could be potentially a decent approach especially when the transformation is exact or incurs low approximation error. More insight on the properties of formulation will be given in next sections.

5.2.3 NLP formulation

Exact formulation: The section shows that the pricewise function formulated in 5-2 can be exactly formulated as an NLP when the sub-functions satisfy these properties:

Property 1:

$$\min_{D \le D_k} \{ R_k(D), R_{k+1}(D) \} = R_k(D) \qquad \forall k \in \{1, \dots, K-1\} \qquad 5\text{-}6$$

Property 2:

$$\min_{D_k \le D} \{ R_k(D), R_{k+1}(D) \} = R_{k+1}(D) \qquad \forall k \in \{1, \dots, K-1\} \qquad 5\text{-}7$$

The following preposition illustrates that a piecewise speed-density relationship can be formulated as the minimum of the $R_k(.)$'s when the properties above are satisfied.

Preposition: For the continuous piecewise function defined in Equation 5-2 and satisfies the Properties 1 and 2:

$$V(D) = \min_{k}\{ R_k(D)\} \quad \text{5-8}$$

Proof: Properties 1 and 2 compare two adjacent sub-functions and can be used to expand the comparison of the sub-function, $R_k(D)$, versus all other sub-functions. In particular, the condition $D \le D_k$ implies $D \le D_k \le D_{k+1}$ and one can conclude from 5-6 that $R_k(D) \le R_{k+1}(D) \le R_{k+2}(D)$. This can be generalized as:

$$\text{for } D \le D_k\text{: } \min_{k\le j\le K}\{ R_j(D)\} = R_k(D) \quad \text{5-9}$$

Similarly from 5-7, it can be concluded that $R_k(D) \le R_{k-1}(D) \le R_{k-2}(D)$ when $D_{k-2} \le D_{k-1} \le D$. This can be generalized as:

$$\text{for } D_{k-1} \le D\text{: } \min_{1\le j\le k}\{ R_j(D)\} = R_k(D) \quad \text{5-10}$$

Combining 5-9 and 5-10:

$$\text{for } D_{k-1} \le D \le D_k\text{: } \min_{1\le j\le K}\{ R_j(D)\} = R_k(D) \quad \text{5-11}$$

Since Equation 5-11 holds for all the regimes, the proof is complete.

The min operator is a nonlinear function, thus the piecewise function was equivalently formulated as an NLP, which can be solved by appropriate nonlinear solvers. Hence the following sets of constraints (Equations 5-12 to 5-14) can be appended to the Constraint 3 to formulate piecewise speed-density functions.

$$V'_1(D_{i,t+1}) = R_1(D_{i,t+1}) \qquad \forall i \in I, \forall t \in T \quad \text{5-12}$$

$$V'_k(D_{i,t+1}) = \min\{ R_k(D_{i,t+1}), V'_{k-1}(D_{i,t+1})\} \quad \forall k \in \{2,3,\dots,K\}, \forall i \in I, \forall t \in T \quad \text{5-13}$$

$$V(D_{i,t+1}) = V'_K(D_{i,t+1}) \qquad \forall i \in I, \forall t \in T \quad \text{5-14}$$

Approximate formulation: When the Properties 1 and 2 are not satisfied for a speed-density relationship, the relationship can be approximated by functions that possess the Properties. The question is if such an approximation exists and what is the form of that approximation. This sections shows that a homographic approximation always exists when the corresponding flow-density model is concave and continues.

Proposition: Consider a flow-density curve (see Figure 5-2) that is approximated by a continues and piecewise linear function L(D)

Such that

$$L(D) = \sum_{k=1}^{K'} (\alpha_k + \beta_k D) * I_k(D) \qquad 5\text{-}15$$

Where,

$I_k(D)$ is the indicator function to determine the sub-function, similarly defined in 5-2

and α_k's and β_k's are constants with the following relationships satisfying concavity of L(D):

$$\alpha_k < \alpha_{k+1}$$

$$\beta_{k+1} < \beta_k$$

Then the corresponding speed-density relationship holds the properties 1 and 2.

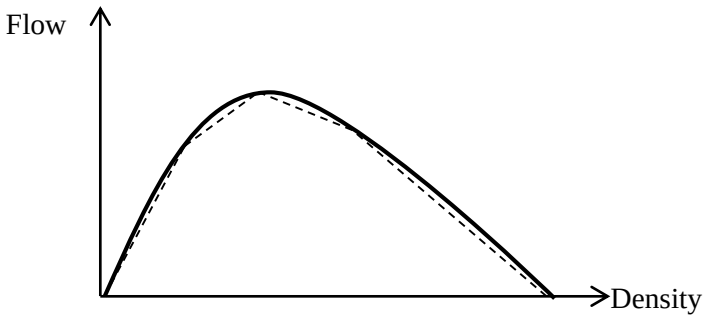


Figure 5-2: Piecewise smooth flow-density relationship (solid curve) approximated by piecewise linear function (dashed lines)

Proof. The corresponding speed-density relationship is composed of homographic functions of the following form:

$$V(D) = \sum_{k=1}^{K} \left(\frac{\alpha_k}{D} + \beta_k\right) * I_k(D)$$

Since the piecewise linear approximation is continuous and $\beta_{k+1} < \beta_k$,

$$L_k \leq L_{k+1} \quad \text{when } D \leq D_k$$

And consequently

$$R_k = \frac{L_k}{D} \leq \frac{L_{k+1}}{D} = R_{k+1} \quad \text{when } D \leq D_k$$

As a result, $\min_{D \leq D_k}\{R_k(D), R_{k+1}(D)\} = R_k(D)$, which expresses the Property 1.

Similarly, one can verify that the Property 2 holds for the approximation. Since the piecewise linear approximation is continuous and $\beta_{k+1} < \beta_k$,

$$L_k \geq L_{k+1} \quad \text{when } D_k \leq D$$

Then

$$R_k = \frac{L_k}{D} \geq \frac{L_{k+1}}{D} = R_{k+1} \quad \text{when } D_k \leq D$$

As a result, $\min_{D_k \leq D}\{R_k(D), R_{k+1}(D)\} = R_{k+1}(D)$, which expresses the Property 2.

5.2.4 Practical consideration

There is a trade-off between the accuracy of homographic approximation and resulting computational effort. To achieve the desirable accuracy one may use more homographic sub-functions, but these results in more constraints, and more nonlinear function evaluations. If computational complexity matters, the study recommends other forms of approximation, such as power function, be used, especially for the oversaturated conditions to reduce the number of sub-functions. The next section illustrates approximation process for HCM 2010 speed-flow relationships and work zone speed-flow relationships developed by Benekohal and Ramezani (2013).

5.3 Speed-density relationship approximation

This section of the paper illustrates examples for the approximation process. The following two sets of relationships are approximated:

1) The HCM 2010 relationships for basic freeway sections, and

2) Work zone relationships developed by Benekohal et al. (12).

These sets describe the relationships between speed and flow, but the WZSH formulation uses the speed-density relationship. Thus, first the speed-flow relationships are approximated

using a few simple functions. Then, the corresponding speed-density relationships are obtained which are used in the WZSH formulation, while they maintain the Properties 1 and 2. It should be noted that the HCM 2010 relationships which is for non-work zone sections of a freeway is included here because speed harmonization may be applied over a long stretch of a freeway that may include both work zone and non-work zone sections of a freeway.

5.3.1 Approximation for HCM 2010 relationships

In HCM 2010, there are 5 Sets of speed-flow curves each representing a free flow speed on basic section of a freeway (the range is from 75 mph to 55 mph). Here, detailed explanations are provided when the free flow speed is 75 mph. A similar approach will be applied for other free flow speeds, and will be presented in this section. Then one can convert the speed-flow relationship into speed-density relationship and incorporate in the WZSH.

The HCM-2010 relationship only consists of two regimes for undersaturated conditions:

$$V = \begin{cases} FFS & F < F_B \\ FFS - a(F - F_B)^2 & F_B \leq F \leq F_C \end{cases} \qquad 5\text{-}16$$

Where

F : Flow rate (pcphpl),

V: Speed (mph),

FFS: Free flow speed and is equal to 75 mph,

a : Model coefficient and is equal to -0.00001107,

F_B: Flow rate at the breakpoint, which is equal to 1000 pcphpl, and

F_C: Flow rate at capacity, which is equal to 2400 pcphpl.

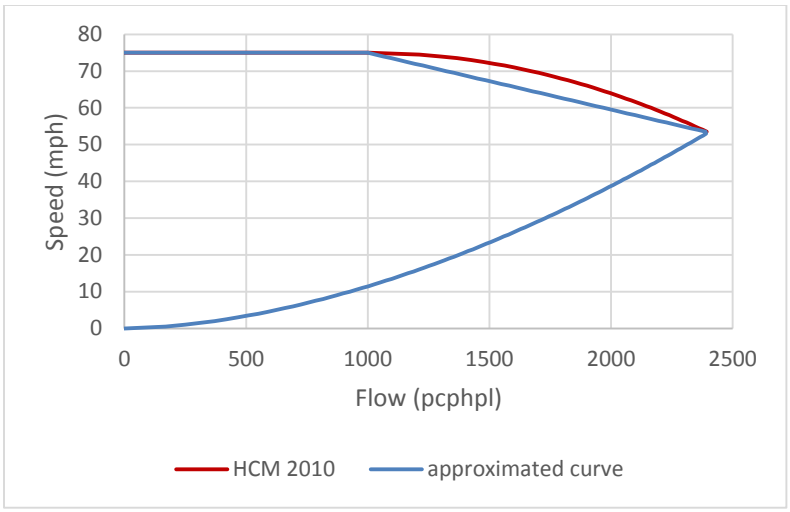
The approximation includes two steps. In Step 1, the transition regime (polynomial equation) is approximated by a line as shown in Figure 5-3. The maximum speed discrepancy between the approximated curve and the HCM curve could be about 5.4 mph with average error of -3.6 mph; however, these discrepancies will be reduced in Step 2, by increasing the break point for the approximated curve. Thus, the formula for the approximated curve with the new break point, $F_B{}'$, will be:

$$V = \begin{cases} FFS & F < F_B{}' \\ FFS + \dfrac{V_C - FFS}{F_C - F_B{}'}(F - F_B{}') & F_B{}' \le F \le F_C \end{cases} \qquad 5\text{-}17$$

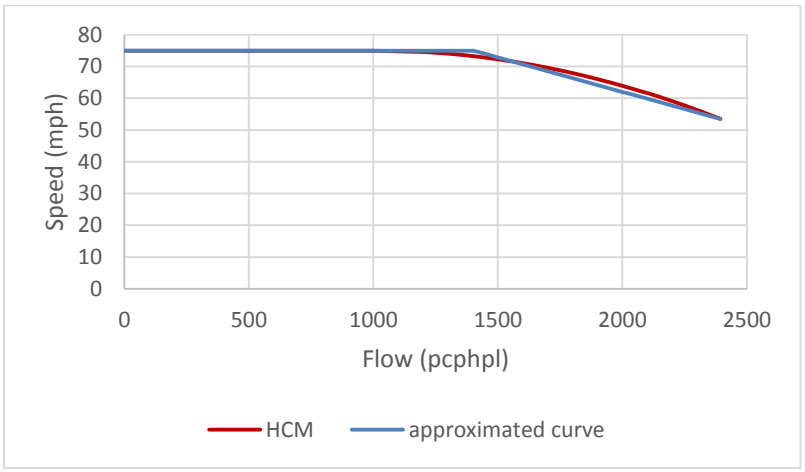
Where,

F_C and V_C are capacity and speed at capacity, respectively.

When $F_B{}' = F_B = 1000$, the approximated curve in Equation 5-17 is the same as that from Step 1 (shown in Figure 5-3a). Increasing $F_B{}'$ can reduce discrepancy as shown in Figure 5-3b for a bending point of 1400 pcphpl. Table 5-1 reports the maximum absolute error and average error for different values of $F_B{}'$ with increments of 100 pcphpl. Errors in the speed estimation were computed over the domain of the transition flow for every 5 pcphpl. The break point, $F_B{}'$, of 1400 pcphpl was selected since it returned the lowest maximum absolute error, 1.94 mph. In addition, the corresponding average error is -0.5 mph which can be acceptable for practical purposes.



Part a



Part b

Figure 5-3: HCM 2010 Speed-Flow Curve Approximation for bending points a) 1000 b) 1400

Table 5-1: Error in Speed Estimation for Different Break Points

F_B'	Maximum Absolute Error (mph)	Average Error (mph)
1000	5.41	-3.59
1100	4.61	-2.82
1200	3.75	-2.05
1300	2.85	-1.27
1400	1.94	-0.50
1500	2.77	0.27
1600	3.99	1.04

Similar approach is applied for the other free flow speeds. Table 5-2 reports the selected F_B values for all the free flow speeds.

Table 5-2: Parameters for the Approximated Speed-Flow Curves for HCM 2010 Speed-Flow Relationships

FFS	Undersaturated part			Oversaturated part	
	F_C (Capacity)	V_C (Speed at Capacity)	F_B' (Break Point)	a	b
75	2400	53.3	1400	250	0.5688
70	2400	53.3	1500	250	0.5688
65	2350	52.2	1600	250	0.5665
60	2300	51.1	1700	250	0.5641
55	2250	50	1800	250	0.5617

Still the HCM-2010 relationships do not have any equations for the oversaturated conditions. Since the purpose of this section is to illustrate the approximation process, some equations can be suggested for the oversaturated conditions satisfying the properties. This form of equation may be examined in future when the congested part of the relationship is developed. Benekohal et al. (12) showed that power functions can sufficiently describe congestion data from work zones. The general form of the power function is

$$F = aV^b \qquad 5\text{-}18$$

Where, a and b are the coefficients of the equation. The same form of the equation is used for basic freeway sections with new coefficients, which are different from those for work zones. The coefficients a and b are found such that the power function connects the capacity point to the jam density point. It is assumed that jam density is 250 pc/mi/ln and occurs at a

speed of 1 mph. This jam density is corresponding to an average spacing of 21 ft. when all vehicles are passenger cars. This assumption will determine the parameter a to be 250. For free flow speed of 75 mph, the capacity is 2400 pcphpl and corresponding speed is 53.3 mph. These numbers will return 0.5688 for the value of b. For other free flow speeds, the capacity, the corresponding speed, and the resulting parameters can be obtained from Table 5-2, and

Then the corresponding speed-density relationship is defined as below:

$$V(D) = \begin{cases} FFS & 0 \le D < D_B \\ \dfrac{FFS - mF_B{}'}{1 - mD} & D_B \le D < D_C \\ \left(\dfrac{D}{a}\right)^{\frac{1}{1-b}} & D_C \le D \le D_{jam} \end{cases} \qquad 5\text{-}19$$

Where,

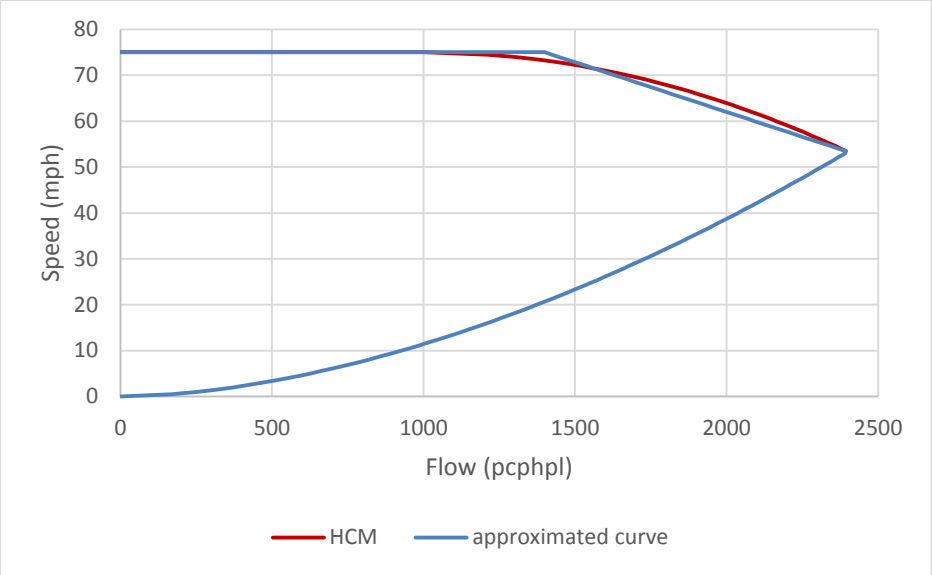
D_B: Density at break point in (pc/mi/ln),

D_C: Density at capacity in (pc/mi/ln),

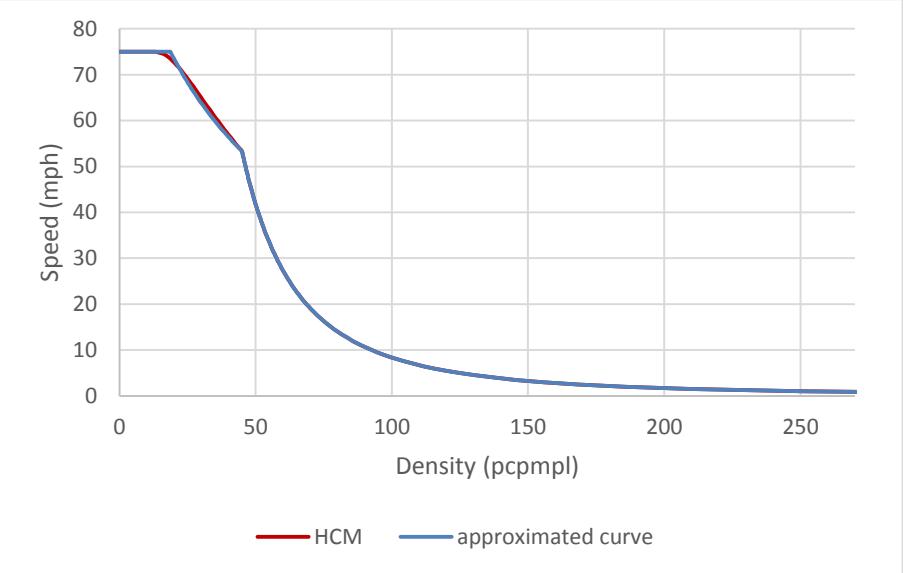
D_{jam}: Jam density in (pc/mi/ln), and

$m = \frac{V_C - FFS}{F_C - F_B{}'}$.

One can easily check that the Properties 1 and 2 are satisfied for the approximated speed-density relationship in Equation 5-19 when density is less than jam density. Figure 5-7-a and Figure 5-7-b show the approximated speed-flow and speed-density curves for free flow speed of 75 mph.



Part a



Part b

Figure 5-4: Curve Approximation a) HCM 2010 Speed-Flow Curves for Speed Limit of 75 mph, b) HCM 2010 Speed-Density Curves for Speed Limit of 75 mph

5.4 Approximation for work zone speed-flow relationships

Benekohal et al. (12) developed four-regime speed-flow relationships for work zones as shown in Figure 5-5. The four regimes are: 1) free flow, 2) upper transition, 3) lower transition,

and 4) congestion. The two transition regimes are four-degree polynomial splines that potentially may create computational difficulties to compute gradient and Hessian matrices. Moreover, there is not always a closed form equation, which strictly defines speed versus density or speed versus flow for these regimes. Thus the curves are approximated with simpler functions. How to approximate the curves are explained for speed limits of 45 mph and 55 mph.

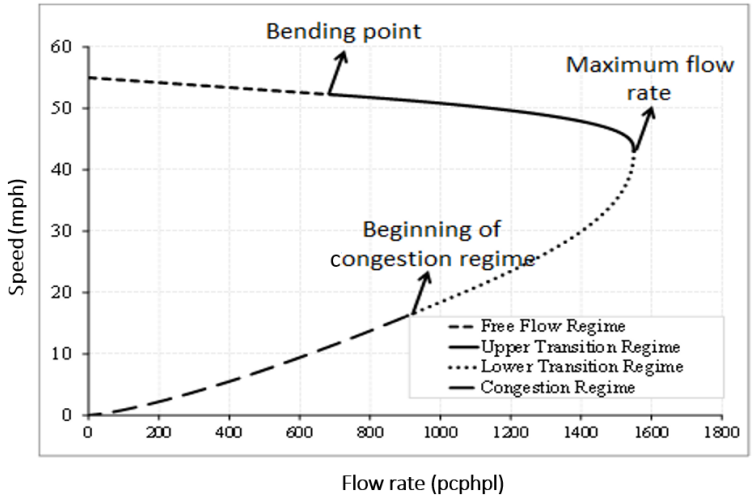


Figure 5-5: General Form of Four-Regime Speed-Flow Relationships

Similar to HCM-2010 curves, the following relationships are used to approximate the undersaturated part of the work zone speed-flow relationships:

$$V(F) = \begin{cases} V_0 & 0 \le F < F_B \\ V_0 + \dfrac{V_C - V_0}{F_C - F_B}(F - F_B) & F_B \le F < F_C \end{cases} \qquad 5\text{-}20$$

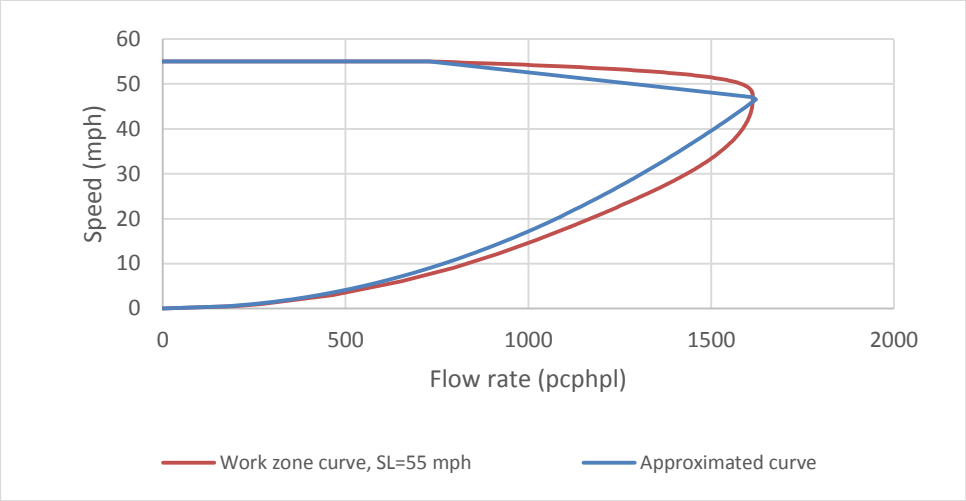
Where V_0 denotes speed at a flow rate of zero. V_0 for the curves with speed limit of 55 mph is equal to the intercept in the original curve providing exact relationship for free flow regime which is a horizontal line in the original curves. However the free flow regime is an inclined line when speed limit is 45 mph. In this case, V_0 is selected to be equal to speed at the break point. This choice of V_0 creates some discrepancies when the traffic volume is very low, but it results in less discrepancy at higher traffic volumes (e.g. zero discrepancy at break point) which are more important for congestion management.

Then for both speed limits, the break point is connected to the capacity point using a line as formulated in the second regime of Equation 5-20.

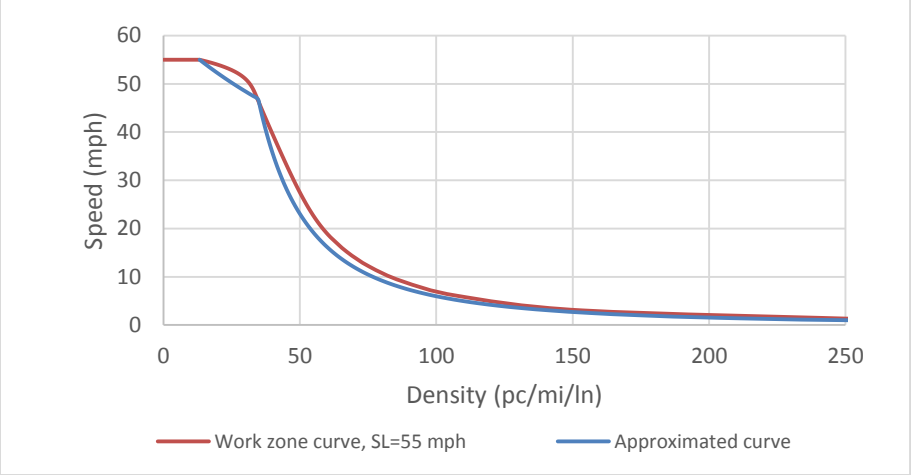
For the oversaturated part, a power function (See Equation 5-18) connects the point with jam density to the capacity point. Similar to the approximations for the HCM-2010 curves, a jam density of 250 pc/mi/ln with a speed of 1 mph is used for the work zone curves. Table 5-2b shows the coefficients of the approximated speed-flow curves for work zones, and Figure 5-7c to Figure 5-7f demonstrate the four-regime curve versus the approximated curve for work zones.

Table 5-3: Parameters for the Approximated Speed-Flow Curves for Work Zone Relationships

	Undersaturated part				**Oversaturated part**	
FFS	**K**	**F_C**	**V_C**	**F_B**	**a**	**B**
55	55.00	1614	47.0	729	250	0.4871
45	42.74	1350	37.7	566	250	0.4681

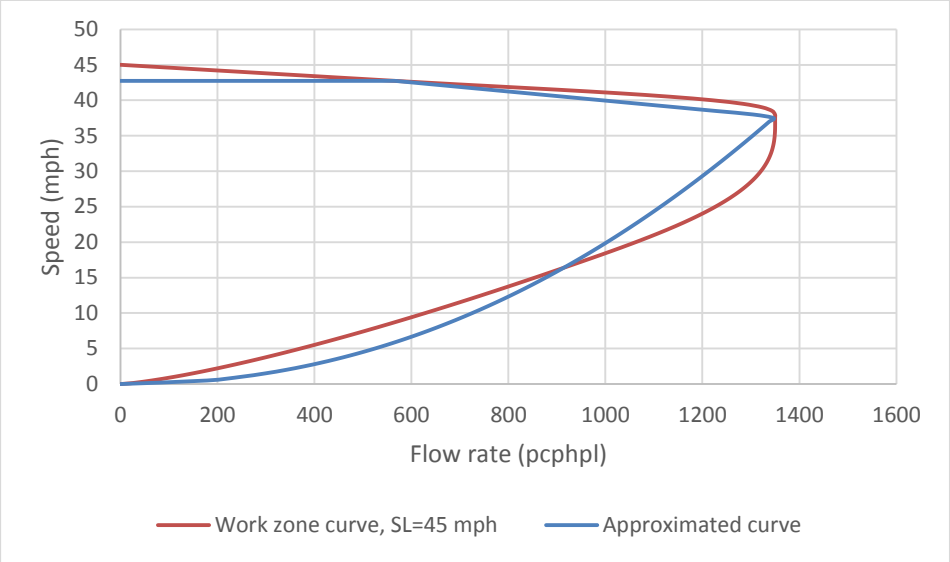


Part a

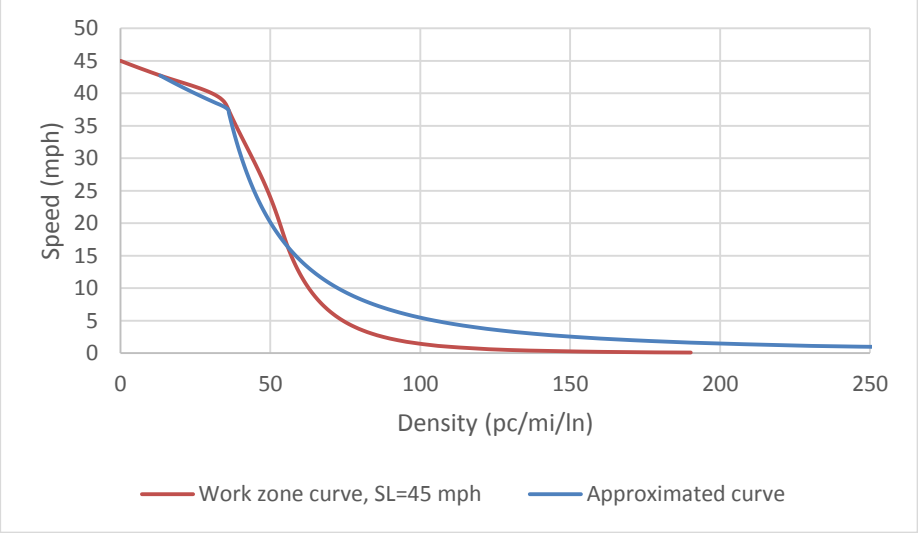


Part b

Figure 5-6: Curve Approximation a) Work Zone Speed-Flow Curves with Speed Limit 55 mph b) Work Zone Speed-Density Curves with Speed Limit 55 mph



Part a



Part b

Figure 5-7: Curve Approximation a) Work Zone Speed-Flow Curves with Speed Limit 45 mph b) Work Zone Speed-Density Curves with Speed Limit 45 mph

5.5 Benchmark problem

A hypothetical benchmark problem is solved to evaluate computational costs and resulting improvement in traffic operation. Advisory speeds are determined for the 12.5-mile roadway shown in Figure 5-8. The roadway is discretized with 0.25-mile long sections. The roadway can be divided into 5 parts as described below.

Part 1) it is a 7-mile basic freeway section with speed limit of 65 mph. This section uses the speed-flow curve proposed by HCM 2010 for basic freeway section with free flow speed of 65 mph. the higher intercept was not selected because it was assumed that compliance rate is 100% when advisory speeds are displayed and this compliance rate is preserved for the static speed limits.

Part 2) the length of this part is a mile and speed limit is 55 mph. In this part, information signs and reduced speed signs are placed due to presence of the work zone. The capacity of this part is 1600 pcphpl. For this part, the speed-flow relationship proposed by Ramezani et al. (2013) for work zones with speed limit of 55 mph and intercept of 55 mph is used. The formulation of this relationship and how to convert to speed-density relationship is illustrated in Section 5.4.

Part 3) the length of this part is 3 miles, speed limit is 55 mph, and one lane is closed. The speed-flow curve used for this section is the same as that used for Part 2.

Part 4) the length of this part is 1 mile and speed limit is further reduced to 45 mph. For this part, the speed-flow relationship (Ramezani et al. 2013) for work zones with speed limit of 45 mph and intercept of 45 mph is used as formulated in Section 5.4. The section 2 in this part, is the work space (indicated by WS in the sketch) where work activity occurs. The capacity of this part is 1350 pcphpl; however there might be some capacity drop in this section depending on the presence and severity of work activity. Part 5) the length of this part is 0.5 miles and it has the same characteristics as Part 1 has. These sections are considered to let vehicles exit the network and make sure that the boundary condition at the exit section does not influence the bottleneck (i.e. Work space)

All the speed-flow curves are approximated and transformed to speed-density relationships as described in Section 5.4.

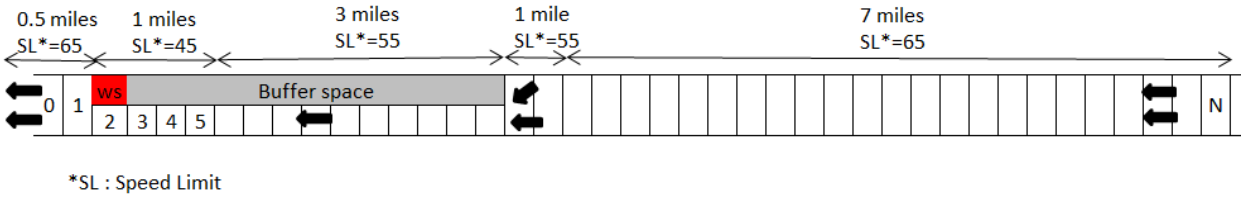


Figure 5-8: Work Zone Sketch for the Example Problem

Also, traffic is composed of only passenger cars and the traffic volume is 1200 pcph. When there is no work activity, capacity of the Work Space is 1350, (determined from the speed-flow relationship) pcphpl which is higher than the traffic volume in the one-lane section. As a result, there is no queue in the roadway at the beginning of analysis; capacity of work space drops from 1350 pcphpl to 950 pcphpl due to severe work activity that lasts for 15 minutes (from Interval 11 to Interval 101). The capacity drop results in a queue that reaches 1.75 miles upstream of the Work Space when there is no speed harmonization.

To control the queue propagation, CMSs are placed in upstream boundary of Sections 2 to 42. A CMS will not be placed in the first two miles of the roadway (sections 43 to 50). Five of the CMSs are critical and the rest of them are non-critical. The first critical CMS is installed 2 miles upstream of the work space. The rest of the critical CMSs are located with a spacing of 2 miles. Hence the critical CMSs are in Sections 10, 18, 26, 34, and 42. The rest of the CMSs are non-critical.

As discussed in Chapter 4, the roadway is discretized by 0.25-mile long sections, and the Δt (i.e. the time interval to update traffic flow variables for each section) is equal to 10 seconds. Optimization time horizon is set to be 300 intervals (=50 min), which is roughly the time that queue dissipates in the Base case (i.e. when there is no speed harmonization) and section densities are below critical density.

The value of second order model parameters are the same as those found in Chapter 4.

Table 5-4: The second order model parameters

Parameter	τ (sec.)	ϑ ($\frac{\text{mile}^2}{\text{h}}$)	κ (pc/mile/ln)	ϕ
Value	27	18	35	2.2

5.5.1 Problem solution

The formulation presented in Chapter 4 assumes that the decision variables are the number of CMS, locations of critical CMSs, and the advisory speeds displayed by CMSs. In this problem, the number and locations of critical CMSs are known and the only decision variables are the advisory speeds. Thus the formulation needs two modifications to solve this problem.

One needs to fix the values of variables $x_{10}, x_{18}, x_{26}, x_{34}, x_{42}$ to 1 and the values of the rest of x_i's to zero, and solves the problem to find advisory speeds.

The problem was solved using LOQO 7.00 (Vanderbei, and Shanno (1999), Vanderbei (1999), which is a solver for large-scale, nonconvex, and nonlinear optimizations. The solver reported that the program includes roughly 158K variables and 186K constraints.

Initial values of the traffic variables, (i.e travel speed, steady state speed, density, and flow rate) are the same as those in Base condition; the initial values of the advisory speeds were set to be equal to maximum of 35 mph and steady state speed in the Base condition.

The maximum iteration was set to be 500; the results showed that more iterations are not needed; as displayed in Figure 5-9 the amount of reduction in the value of objective function was less than 1% over the last 100 iterations (i.e from iteration 400 to 500). The problem was submitted in NEOS server (Czyzyk et al. (1998), Dolan (2001)) and solved in 1372 seconds (≈ 23 minutes). Figure 5-9 shows how the objective function and infeasibility measure change over iterations.

The infeasibility measure was defined in (Vanderbei (1999)) and is explained when there is only one constraint the optimization program and then generalized for more constraints. Once the slack variable is added to an inequality constraint and it becomes equality, the amount of infeasibility is defined as the difference between write hand side and left hand side of the equality. For instance, if slack variable of w is added to the constraint $b \leq ax$, the resulting equality becomes $ax - w = b$. the amount of infeasibility for this constraint is defined as $\rho = b - ax + w$ and the infeasibility measure is computed as:

$$\text{Infeasibility measure} = \frac{\sqrt{\| \rho \|^2}}{\sqrt{\|b\|^2} + 1}$$

For more than one constraint, the infeasibility measure can be interpreted as the second norm of the infeasibility vector divided by the second norm of the vector of constants plus one.

Before iteration 100, there is considerable amount of infeasibility. During these iterations the solver tries to reduce infeasibility while it improves the objective function. The infeasibility is further reduced between iterations 100 and 200. After iteration 200 the infeasibility measure varies between 1.2 *10^-6 and 4.8*10^-6. The objective function in iteration 1 was 234.4 hr,

which is equal to total travel time for the Base condition. After 500 iterations, the objective function is reduced to 227.5 hr.

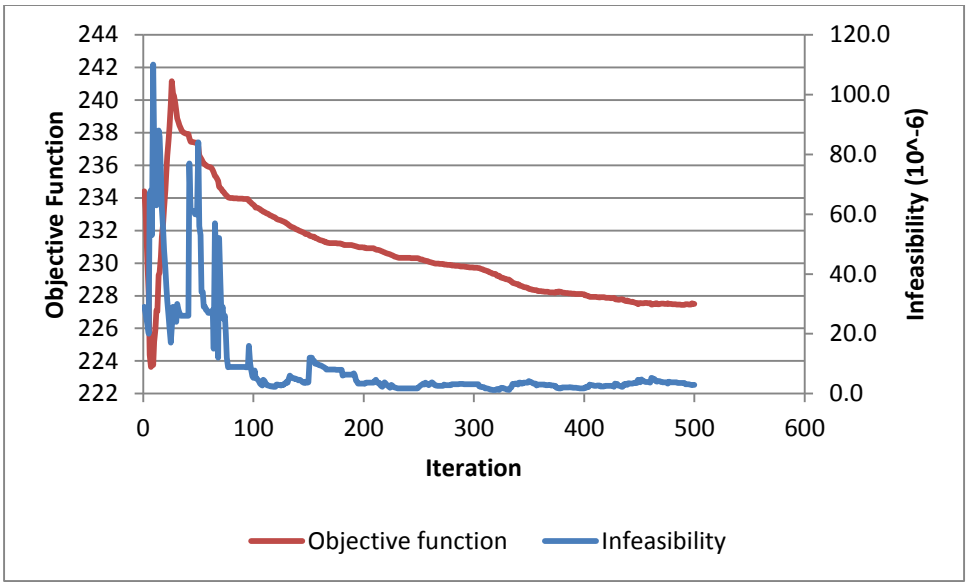


Figure 5-9: Variation of objective functions and infeasibility measure over iterations

5.5.2 Results

LOQO determines the optimal advisory speeds that are decimal as shown in Figure 5-10. The advisory speeds for CMS10 and CMS18 that are closest to the bottleneck were always more than traffic speed to provide maximum throughput. Thus they are not shown in Figure 5-10. The CMS 26, CMS 34 and CMS 42 reduce traffic speed in order to decrease both arrival rate and queue propagation rate in the one-lane section. These CMSs start reducing traffic speed roughly from the beginning of the capacity drop (interval 6) until close to Interval 150. After this interval, these CMSs gradually increase the advisory speeds and display values close to the static speed limit of 65 mph.

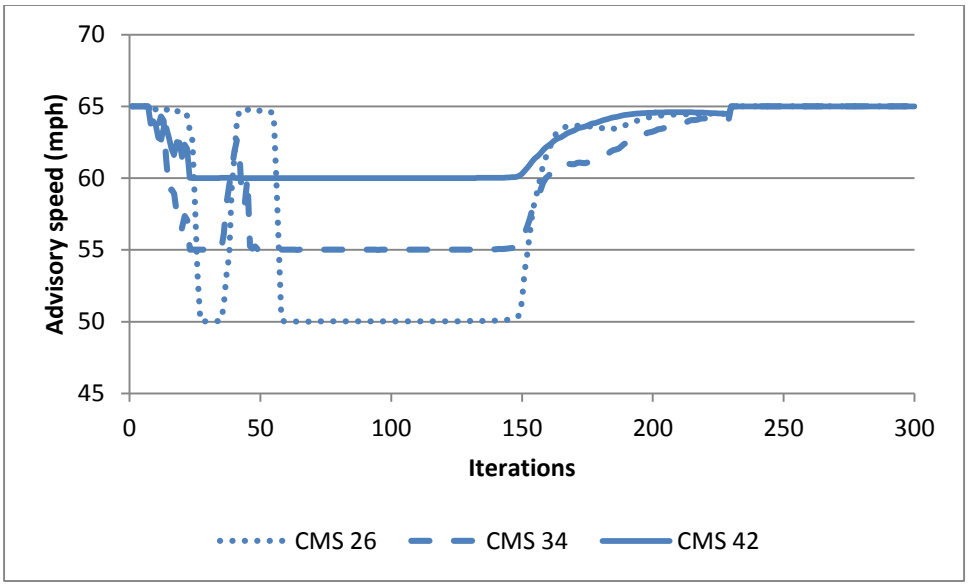


Figure 5-10: Decimal advisory speeds

In real world advisory speeds are multiple of 5; thus the decimal advisory speeds are rounded to the nearest multiple of 5 under three scenarios: 1) round up to the nearest multiple of 5, denoted by R5+, 2) regular rounding to the nearest multiple of 5, denoted by R, and 3) round down to the nearest multiple of 5, denoted by R5-. The Table 5-5 shows the resulting total travel time, total delay, and percentage of delay reduction.

The total delay is computed as:

$$\text{Total delay} = \text{TTT} - \text{HVOL} * N_t * \Delta t * TT_{SL} \qquad 5\text{-}21$$

Where, TTT is the total travel time. The second term in Equation 5-21 is equal to the total travel time under the posted speed limits. The definitions of variables in the second term are:

HVOL : Hourly volume in pcph,

N_t. Number of time intervals in the study period,

Δt: Interval length in hr, and

TT_{SL}: Travel time through the control sections under the posted speed limits.

The decimal advisory speeds reduce total delay by 18.3%. When advisory speeds are rounded to the nearest multiple of 5, speed harmonization reduces total delay by 15.7%. This is the closest value to the percentage of delay reduction obtained by decimal advisory speeds. If the advisory speeds are rounded down (R5-) the percentage of delay reduction is 8% which is considerably less than regular rounding scenario. If the advisory speeds are rounded up (R5+),

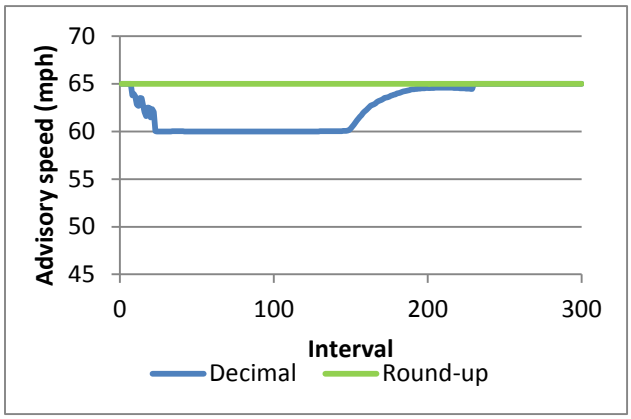
the speed harmonization deteriorates the traffic operation and increases delay by 3% compared with the Base condition.

Table 5-5: Travel time and percentage of delay reduction for different forms of displayed advisory speeds

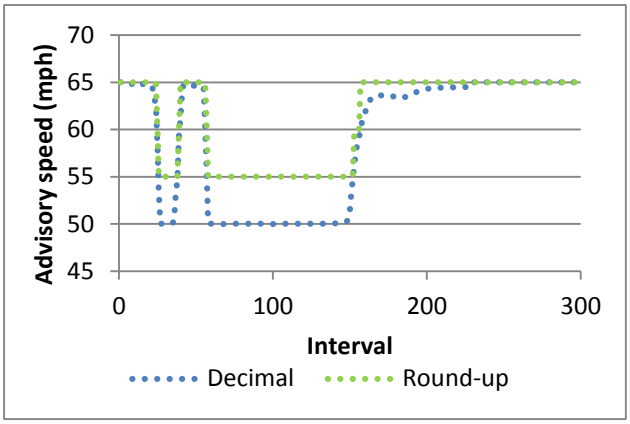
Display advisory speed form	**Total travel time (hr)**	**Total delay (hr)**	**Total delay in base condition**	**Percentage of delay reduction**
Fractional	227.5	30.8	37.7	18.3
Round down to multiple of 5 (R5-)	231.4	34.7		8.0
Round to nearest multiple of 5 (R)	228.5	31.8		15.7
Round up to multiple of 5 (R5+)	235.5	38.8		-3.0

To further investigate the effect of the three scenarios for rounding, the resulting advisory speeds are compared with the decimal advisory speeds for CMSs 26 and 42 that are respectively the most downstream and most upstream effective CMSs based on the decimal advisory speeds.

As shown in Figure 5-11, when advisory speeds are rounded up, CMS 42 (most upstream CMS) display 65 mph which is equal to the static speed limit. This means that this CMS does not have any influence on traffic. Thus in general rounding up could practically reduce the number of critical CMSs. Also for other CMSs, the displayed speeds are greater than when decimal advisory speeds are displayed. For instance, for CMS 26, the minimum rounded-up advisory speed is 55 mph while the corresponding decimal value is 50.0114 mph. This could cause higher shockwave propagation because a greater advisory speed resulted in a higher delay. Thus in R5+ scenario it is possible to practically deactivate one of the critical CMSs. In addition, it may impose reduced speed to traffic, while this speed reduction does not help to decrease shockwave propagation speed. For example, in this benchmark problem, this scenario causes 3% extra delay compared with Based conditions.



Part a

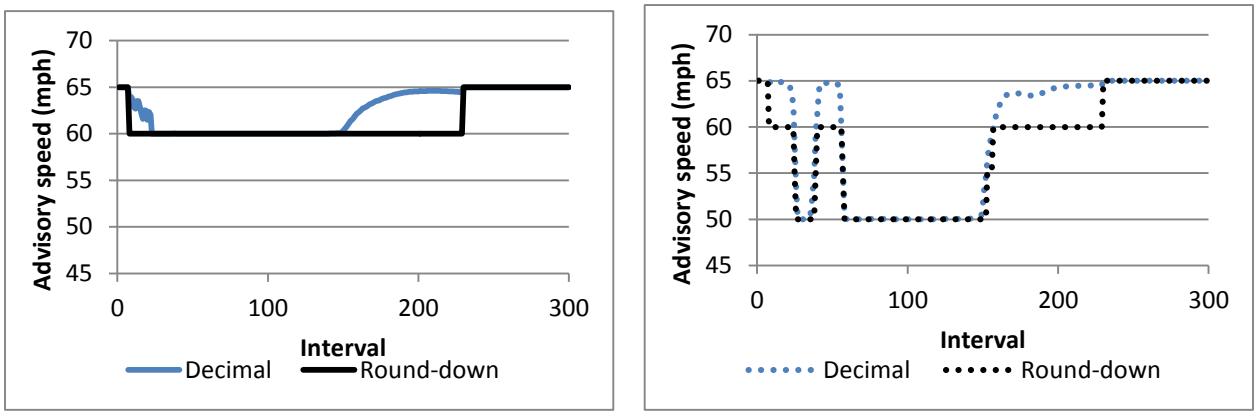


Part b

Figure 5-11: Decimal advisory speeds versus rounded up advisory speeds to the nearest multiple of five a) CMS 42, b) CMS26

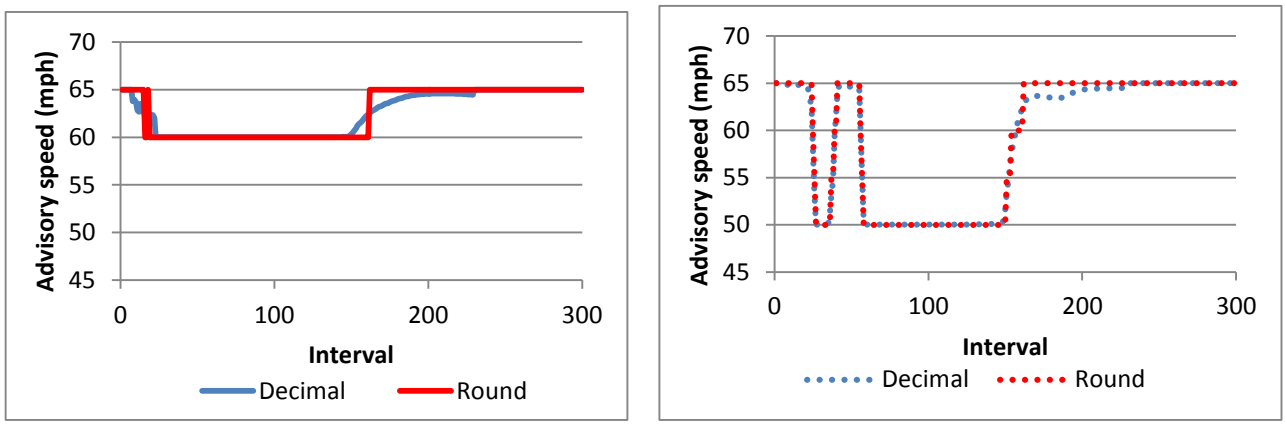
When advisory speeds are rounded down, there could be extra speed reduction while there are no needs. This is more significant when decimal advisory speeds are changing from a high value to a low value or vice versa. For example in Figure 5-12, during intervals 8 to 22 both CMSs 26 and 42 try to gradually reduce traffic speed by 5 mph when decimal advisory speeds are displayed; however this change is abrupt and happens at interval 8 when advisory speeds are rounded down. This means that if advisory speeds are rounded down, the traffic speed may be reduced earlier and in higher amount than needed. Also, during intervals 157 to 229 both CMSs display decimal advisory speeds that are close to static speed limit of 65 to let traffic gain speed and move faster. When advisory speeds are rounded down, traffic speed is kept to be 5 mph below speed limit while there is no queue in the field and no need for reduced traffic speed. This continues until interval 229 after which advisory speed of 65 is displayed. As a result, the (R5-) scenario resulted in 8% delay reduction which is considerably less than 18.3% which is the percentage of delay reduction when the decimal advisory speeds are displayed.

When advisory speeds are rounded to the nearest multiple of 5, the resulting advisory speeds are more similar to decimal values than the other two forms of rounding as shown in Figure 5-13. This also results in 15 % of delay reduction which is close to 18.3% delay reduction achieved by decimal advisory speeds.



Part a Part b

Figure 5-12: Decimal advisory speeds versus rounded down advisory speeds to the nearest multiple of five a) CMS 42, b) CMS26



Part a Part b

Figure 5-13: Decimal advisory speeds versus rounded advisory speeds to the nearest multiple of five a) CMS 42, b) CMS26

The rounded advisory speeds are shown in Figure 5-14 for the critical CMSs. The CMS26, CMS34 and CMS42 reduce traffic speed in order to decrease both arrival rate and queue propagation rate in the one-lane section. In particular, roughly after 2 minutes from the beginning of the capacity drop (around Interval 14), these CMSs start reducing traffic speed until close to Interval 162. After this interval, these CMSs display the static speed limit of 65 mph.

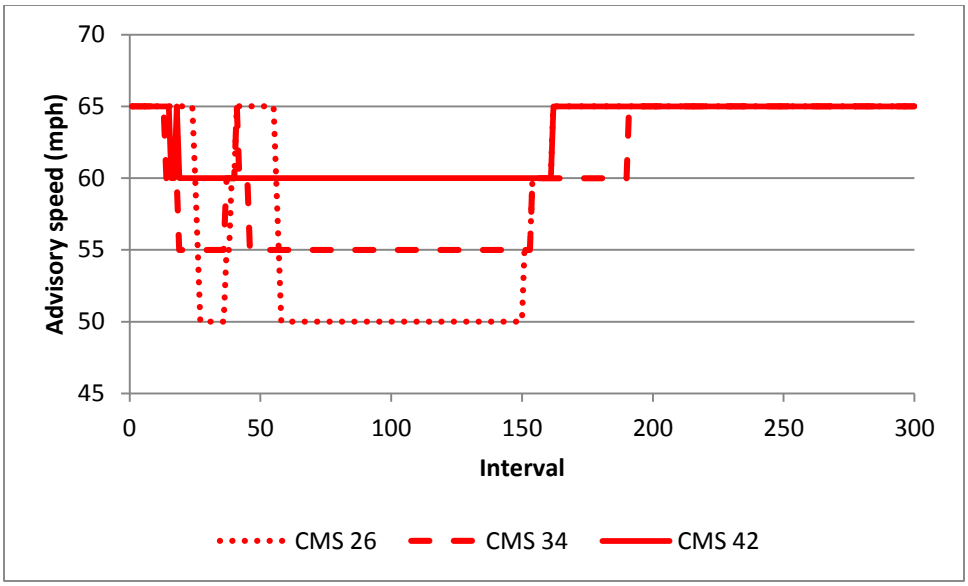
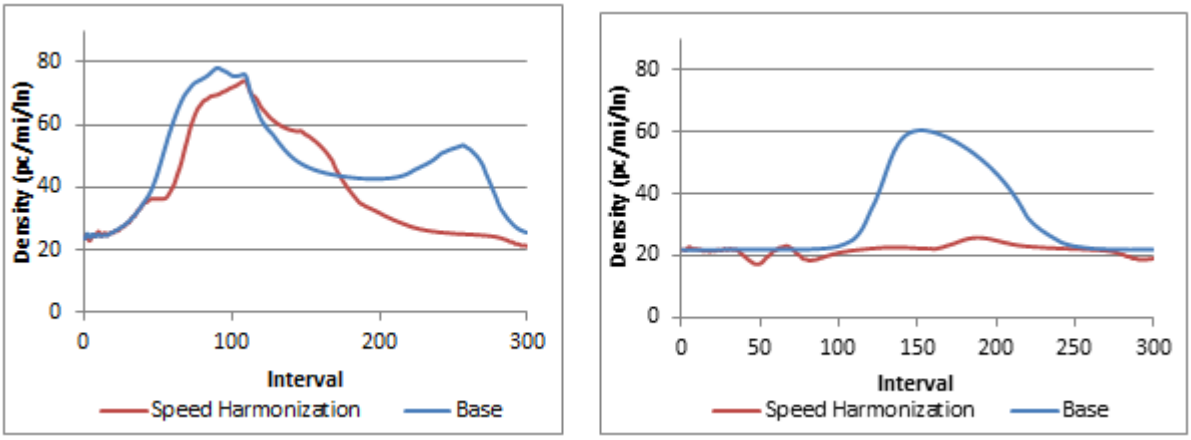


Figure 5-14: Rounded advisory speeds

To explore how the WZSH influences congestion propagation, Figure 5-15 display density (pc/mi/ln) variations for Sections 5 and 9 that are 1 mile and 2 miles away from the end of work space, respectively. Density at capacity is 42 pc/mi/ln, thus these sections experience congestion when their density exceeds 42. In the Base case queue reaches to section 9 while this section does not experience queuing condition when there is speed harmonization. When there is speed harmonization, Section 4 experiences congestion 2 minutes longer than Base conditions. Also duration that this section is in congested is reduced from 28 minutes to 19 minutes.

Overall the speed harmonization reduced delay by 15.7%. Maximum queue length reduced from 2 miles to 1.25 miles which is corresponding to 37.5% reduction. Congestion duration reduced from 47 minutes to 31 minutes which is corresponding to 33% reduction.



Part a Part b

Figure 5-15: Density changes a) 1 mile b) 2 miles from end of work space

5.6 Conclusions

Optimized speed harmonization problem with piecewise speed-density functions is hard and time consuming to solve, because of presence of nonlinear constraints and integer variables. This chapter developed a continuous transformation methodology that eliminates all integer variables that are defined due to the piecewise functions. It was shown that the number of integer variables that are eliminated is proportional to 1) the number of roadway sections, 2) number of time intervals, and 3) number of sub-functions (in the speed-density function) minus 1.

The methodology was used to solve speed harmonization problem for a 12-mile roadway and a 50-min analysis duration. For this problem, the methodology eliminated 28800 integer variables (there are 48 sections, 300 time intervals, and 3 sub-functions) and solved the problem in 23 minutes when the program includes roughly 158K variables and 186K constraints. The results showed that the speed harmonization can reduce delay by 15.7% and max queue length by 37.5%. Thus instead of solving a mixed integer nonlinear program, which is very challenging to solve for large scale problem, one can solve the corresponding nonlinear program and significantly reduce computational difficulties. The transformation methodology provides the opportunity to use more realistic speed-density relationships, rather than very simplified models, and potentially enhances the reliability of speed harmonization techniques for field implementation.

The methodology relies on some mathematical properties of the sub-functions, and it does not depend on the shape of the objective function or other constraints; as a result, one may

apply the methodology for both minimization and maximization problems with different objective functions. It was shown that when the properties 5-6 and 5-7 do not exist in a speed-density relationship, one can approximate the relationship using a piecewise function that satisfies the properties. The approximate transformation always exists when the corresponding density-flow relationship is concave and continuous. The approximation process was illustrated for some piecewise traffic stream models that have been used for practical traffic analysis.

It should be noted that this methodology only eliminated integer variables due to piecewise speed-density relationships; there could be additional integer variables in the formulation when the number and locations of signs are decision variables. Next chapter presents methodologies to solve the mixed integer nonlinear program when the source of integrality is the location and number of signs.

Chapter 6 Sign location problem for optimized speed harmonization

As explained in Chapter 5, there are two types of Changeable Message Signs (CMSs) that can be used for speed harmonization: Critical CMSs and non-critical CMSs. Critical CMSs can display an advisory speed different from the advisory speed displayed by the immediate upstream CMS. Non-critical CMSs always show the advisory speed displayed by the immediate upstream CMS. The number of critical CMSs can influence the effectiveness of speed harmonization. When the number of critical CMSs is insufficient there may not be enough opportunity to influence traffic speed and reduce shockwave propagation speed. As a result, insufficient number of CMSs may cause significantly less-than-optimal delay reduction. On the other hand, too many signs may cause extra maintenance costs and will result in an inefficient design.

Determining the number of signs and their locations are interrelated. In theory, minimum number of signs are used when they are installed in proper locations. As will be shown in the benchmark problem (Section 6.2.1.1), computations show that there are certain locations that CMSs are not effective at all; and there are some locations that CMSs are effective but significantly below the optimal level. Thus, a problem needs to be solved to determine both optimal number and locations of CMSs to minimize travel cost and any other costs such as maintenance costs. This problem is called CMS-Location Problem (CMS-LP).

The current chapter focuses on developing solution methodologies to solve CMS-LP. First, a complexity analysis is performed to discuss challenges in solving the problem. Then three methodologies to solve the problem are proposed. A benchmark problem is solved using the methodologies and the results and computational efforts are compared. At the end, recommendations on how to select a methodology are presented.

6.1 Analysis of computational complexity

Challenges for solving CMS-LP stem from two sources: 1) size of the problem, and 2) presence of integer variables.

The CMS-LP is a large scale problem. This is because the number and location of the CMSs should be optimized over the entire analysis period; As a result, the optimization program, formulated in chapter 3, includes many variables and constraints. To clarify this, Table 6-1

counts the number of variables and constraints where N_t and N_s denote the number of time intervals and the number of sections, respectively.

The binary variable x_i determines if a sign should be placed in section i or not. This variable does not change over time and only changes over space. Thus the number of variable x_i is equal to the number of sections (N_s). Another variable is advisory speed, $AS_{i,t}$, that changes over time and space and the program includes $N_s * N_t$ of these variables. Similarly, the variables density ($D_{i,t}$), exit flow rate ($F_{i,t}$), steady state speed ($U_{i,t}$), speed computed from speed-flow curve ($V(D_{i,t})$), and speed computed from the second order model ($V_{i,t}$) change over time and space and the number of each variable is $N_s * N_t$. As a result, total number of variables is $6 * N_s * N_t + N_s$. The main factor that influences the number of variables is spatiotemporal size of the problem; thus this term is considered to establish a lower bound which is $6 * N_s * N_t$.

Constraints 1 to 8 are defined for each time interval and section, so the number of each constraint is proportional to $N_s * N_t$. Constraints 5 is composed of two inequalities, thus there are $2 * N_s * N_t$ of this constraint. Similarly, Constraint 6 is composed of to two inequalities ($(AS_{i,t+1} - AS_{i,t} \leq \Delta_t)$ and $(-\Delta_t \leq AS_{i,t+1} - AS_{i,t})$) resulting $2 * N_s * N_t$ constraints. There are also $2*N_t$ constraints for boundary conditions and $6*N_s$ constraints for initial conditions. The counting is performed over the constraint sets, as reported in Table 6-1, and the total number of constraints will be $10 * N_s * N_t + 2 * N_t + 7 * N_s - m + 1$. The spatiotemporal size of the problem is the main factor to determine the number of constraints. Hence a lower bound of $10 * N_s * N_t$ is considered for the number of constraints. It should be noted that the constraints for the sign of the variables are not counted in this enumeration since they are simple and will not result in significant computations.

The lower bounds are computed for one-hour analysis of a 25-mile roadway. Given, an hour of analysis period and $\Delta t = 10$ sec, the value of N_t will be equal to 360. If the roadway length is 25 miles and $L = 0.25$ miles, the N_s will be equal to 100. For this problem, the lower bound for the number of variables is 216,000 variables ($=6 * N_s * N_t$) and the lower bound for the number of constraints is 360,000($= 10 * N_s * N_t$).

Thus the problem can be considered as a large scale problem due to the large number of variables and constraints. The challenging aspect of solving the problem is inclusion of all these

variables and constraints in a single optimization program. This is because the optimal number and location of CMSs have to be determined for the entire analysis interval.

Another solution challenge is presence of integer variables. As illustrated in the problem formulation, the problem includes integer variables x_is to determine CMSs locations and they are combined with nonlinear constraints (e.g. Constraint 2). This problem is a Mixed Integer Nonlinear Program (MINLP) which is difficult to solve for a large scale problem.

In this chapter, methodologies are developed to solve the problem and handle integrality and size of the problem.

Table 6-1: Variable and constraint counts

Variables	**Variable Set**	**Number**
x_i	$\forall i \in I$	N_s
$AS_{i,t}$, $V_{i,t}$, $D_{i,t}$, $F_{i,t}$, $V(D_{i,t})$, $U_{i,t}$	$\forall i \in I, \forall t \in T$	$6 * N_s * N_t$
Constraints	**Constraints Set**	**number**
Constraint 1) $D_{i,t+1} * L = D_{i,t} * L + \left(F_{i+1,t} * \lambda_{i+1} - F_{i,t} * \lambda_i\right) * \Delta t$	$\forall i \in I, \forall t \in T$	$N_s * N_t$
Constraint 2) $V_{i,t+1} = \text{Max}\{0, V_{i,t} + \frac{\Delta t}{L} V_{i,t}\left(V_{i+1,t} - V_{i,t}\right) + \frac{\Delta t}{\tau}\left(U_{i,t} - V_{i,t}\right) - \frac{\Delta t * \vartheta}{\tau * L}\left(\frac{D_{i-1,t} - D_{i,t}}{D_{i,t} + \kappa}\right)\}$	$\forall i \in I, \forall t \in T$	$N_s * N_t$
Constraint 3) $U_{i,t+1} = \min\{V(D_{i,t+1}), AS_{i,t+1}\}$	$\forall i \in I, \forall t \in T$	$N_s * N_t$
Constraint 4) $F_{i,t+1} = D_{i,t+1} * V_{i,t+1}$	$\forall i \in I, \forall t \in T$	$N_s * N_t$
Constraint 5) $-\text{Max}AS_i x_i \leq \left(AS_{i+1,t} - AS_{i,t}\right) \leq \Delta_s x_i$	$\forall i \in I, \forall t \in T$	$2 * N_s * N_t$
Constraint 6) $\lvert AS_{i,t+1} - AS_{i,t} \rvert \leq \Delta_t$	$\forall i \in I, \forall t \in T$	$2 * N_s * N_t$
Constraint 7) $AS_{i,t} \leq \text{Max}AS_i$	$\forall i \in I, \forall t \in T$	$N_s * N_t$
Constraint 8) $\text{Min}AS_i \leq AS_{i,t}$	$\forall i \in I, \forall t \in T$	$N_s * N_t$
Constraint 9) $\sum_{i \in I} x_i = n$	-	1
Constraint 10) $\sum_{i=j}^{m+j-1} x_i \leq 1$	$j \in \{1, \ldots, (N_i - m + 1)\}$	N_s-m
Boundary Conditions) $F_{N,t} =$ arrivng volume $D_{0,t} = D_{1,t}$	$\forall t \in T$	$2 * N_t$
Initial Conditions) $V_{i,0}$, $D_{i,0}$, $F_{i,0}$, $V(D_{i,0})$, $U_{i,0}$ are given and are corresponding to arriving volume $AS_{i,0}$=Static speed limit at section i	$\forall i \in I$	$6 * N_s$

6.2 Solution methodologies

Three solution methods are developed such that they progressively reduce computational efforts. The methods are:

1) Greedy Algorithm
2) Augmented-Cut and Branch (AC&B)
3) Approximate Decomposition

The methods will be illustrated and then implemented in a benchmark problem. At the end, the study compares the methods and provides recommendations to use.

6.2.1 Method 1: Greedy Algorithm

The Greedy Algorithm involves more enumeration and determines the solution based on a comprehensive investigation of the search space. The Greedy Algorithm is an iterative method and in each iteration, the location of one critical sign is determined in the control zone. The control zone is defined as the set of the sections where critical CMSs can be potentially located. It is assumed that the control zone includes all sections in the work zone and upstream.

In the first iteration, the algorithm looks for the location of the first sign that would return the minimum value for the objective function. In particular, a sign is located in a given section and the optimization program is solved to determine optimal advisory speeds and the corresponding objective function. Then, the sign is placed in other sections and corresponding objective function values are obtained. The computation that yields the lowest objective function value is selected as the permanent location for the first sign.

In the second iteration, the location of the second sign is found knowing the location of the first sign from the first iteration. The algorithm puts a sign on all possible locations and evaluates the corresponding objective function values. The sections that are evaluated for the second signs must satisfy the minimum spacing constraint (Constraint 10). At the end of the iteration 2, the location of the second sign that would result in the minimum objective function is determined.

Similarly in other iterations, the locations of the remaining signs are found until the additional signs do not improve the objective function, or a stopping criterion is met. The algorithm is presented as below:

Greedy Algorithm Steps:

Define:

I: Set of all section indices,

I_c: Set of section indices which are candidate for permanent installation of a CMS,

I_p^*: Set of section indices which have been permanently selected for installation of a CMS,

I^*: Set of section indices with a critical CMS to evaluate. This is the union of sections with permanent signs (i.e. I_p^*) , and a section $q \in I_c$ that is a candidate for permanent selection,

n: Iteration counter,

m: Minimum number of sections between two consecutive critical CMSs, and

P: The problem formulated in chapter 3 to find the critical CMSs locations.

Step 0: $I_p^* := \emptyset$, n:=1, $I_c := I$, $F_{min}^{(0)}$=total travel time in Base condition

Step 1: $\forall q \in I_c$:

$I^* := I_p^* \cup \{q\}$ and solve P when

$$x_i = \begin{cases} 1 & \text{if } i \in I^* \\ 0 & \text{Otherwise} \end{cases}$$

Step 2: Determine section $q_{min}^{(n)}$ which resulted in the minimum objective function in step 1 and let $F_{min}^{(n)}$ be the corresponding objective function

Step 3: Update I_p^*:

$$I_p^* := I_p^* \cup \{q_{min}^{(n)}\}$$

Step 4: Update I_c:

$$I_c := \{q | q \in I_c \text{ and } |q - q_{min}| \geq m\}$$

Step 5: If $F_{min}^{(n)} - F_{min}^{(n-1)} > 0$ then stop; otherwise n:=n+1 and go to step 1.

It should be noted that the objective function formulated in chapter 3 is

$$\text{Min}((\sum_i \sum_t D_{i,t} * \lambda_i * L * \Delta t) + P(n) \qquad 6\text{-}1$$

In which the first term $(\sum_i \sum_t D_{i,t} * \lambda_i * L * \Delta t)$ is the total travel time and the second term $(P(n))$ is the penalty function for the number of signs. The study uses the following linear penalty function

$$P(n) = p * n$$

Where,

p is the marginal cost for installation of one sign. The marginal cost, p, is defined as the average free flow travel time over the entire control zone, computed as:

$$\text{Avg. free flow travel time} = \sum_i \frac{L}{SL_i} \quad 6\text{-}2$$

Where,

SL_i : the posted speed limit (mph) effective in section i

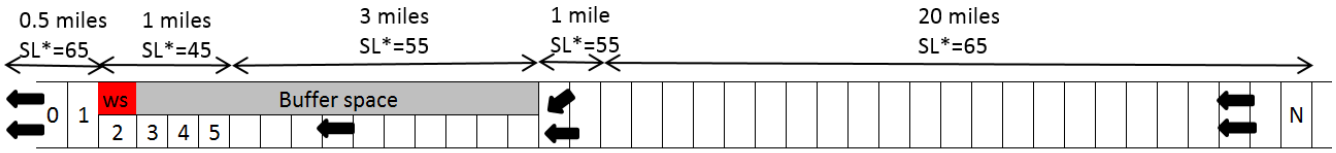
This means that an extra sign is placed if the resulting reduction in travel time is at least equal to the average free flow travel time. The average free flow travel time was selected because it is independent of volume and available capacity. Alternatively one could use average or a portion of travel time in Base condition (i.e. when there is no speed harmonization) but it is a function of volume or capacity in the Base condition and can change from one day to another day. It should be noted the unit of the penalty and the total travel time should be in hour.

6.2.1.1 Benchmark Problem

A hypothetical benchmark problem is solved to illustrate the algorithm implementation and to evaluate its computational cost and resulting improvement in traffic operation. The Greedy Algorithm will be implemented for a 25.5-mile roadway shown in Figure 6-1. The roadway is divided into 5 parts as described below:

Part 1) it is a 20-mile basic freeway section with speed limit of 65 mph. For this section, the study uses the speed-flow curve proposed by HCM 2010 for basic freeway section with free flow speed of 65 mph. The higher intercept was not selected because it was assumed that compliance rate is 100% when advisory speeds are displayed and this compliance rate is preserved for the static speed limits as well. The capacity of this section is 2350 pcphpl. Section 5.4 illustrates how to formulate the corresponding speed-density relationship.

Part 2) the length of this part is a mile and speed limit is 55 mph. In this part, information signs and reduced speed signs are placed due to presence of the work zone. The capacity of this part is 1600 pcphpl. For this part, the speed-flow relationship proposed by Benekohal et al. (2010) for work zones with speed limit of 55 mph and intercept of 55 mph is used. The formulation of this relationship and how to convert to speed-density relationship is illustrated in Section 5.4.



**SL=Speed limit*
Drawing not to scale

Figure 6-1: Roadway information for the benchmark problem

Part 3) the length of this part is 3 miles; speed limit is 55 mph; and one lane is closed. The speed-flow relationship used for this section is the same as that for Part 2.

Part 4) the length of this part is 1 mile and speed limit is further reduced to 45 mph. For this part, the speed-flow relationship (Benekohal et al. 2013) for work zones with speed limit of 45 mph and intercept of 45 mph is used as formulated in Section 5.4. The section 2 in this part, is the work space (indicated by WS in the sketch) where work activity occurs. The capacity of this part is 1350 pcphpl; however there might be some capacity drop in this section depending on the presence and severity of work activity.

Part 5) the length of this part is 0.5 miles and its characteristics is the same as Part 1. These sections are considered to let vehicles exit the network and make sure that boundary conditions at the exit section does not influence the bottleneck (i.e. Work space)

In this problem, the traffic volume is 1200 pcph and is composed of only passenger cars. When there is no work activity, capacity of the Work Space is 1350 pcphpl which is higher than the traffic volume in the one-lane section. As a result, there is no queue in the roadway at the beginning of analysis; capacity of Work Space drops from 1350 pcphpl to 950 pcphpl due to severe work activity that lasts for 15 minutes. The capacity drop results in a queue that reaches 1.75 miles upstream of the Work Space when there is no speed harmonization.

The roadway is discretized by 0.25-mile long sections, and the Δt (i.e. the time interval to update traffic flow variables for each section) is equal to 10 seconds. The optimization time horizon is set to be 300 intervals (=50 min), which is roughly the time that queue dissipates in the Base case (i.e. when there is no speed harmonization).

The value of the second order model parameters are the same as those found in Chapter 4.

Table 6-2: The second order model parameters

Parameter	τ (sec.)	ϑ ($\frac{\text{mile}^2}{\text{h}}$)	κ(pc/mile/ln)	φ
Value	27	18	35	2.2

Based on the objective function, introduced in Equation 6-1, a marginal penalty of 0.403 hr will be considered for each critical sign. The marginal penalty is equal to the travel time for parts 2 to 5 under the posted speed limits. This was simply compute as:

$$\text{Avg. free flow travel time} = \frac{1}{45} + \frac{3}{55} + \frac{1}{55} + \frac{20}{65} = 0.403 \text{ hr.} \quad 6\text{-}3$$

The part 1 was not included in this calculation because it is downstream of the bottleneck and is not in the control zone. Thus the objective function becomes

$$\text{Objective function} = \text{Total travel time} + 0.403 * \text{no. of critical signs} \quad 6\text{-}4$$

$$= (\sum_{i}\sum_{t} D_{i,t} * \lambda_i * L * \Delta t) + 0.403 * n$$

The effect of speed harmonization and sign locations on delay is evaluated as well. The total delay is computed as:

$$\text{Total delay} = TTT - HVOL * N_t * \Delta t * TT_{SL} \quad 6\text{-}5$$

Where,

TTT is the total travel time which is equal to the first term $(\sum_i \sum_t D_{i,t} * \lambda_i * L * \Delta t)$ in the objective function.

HVOL : Hourly volume in pcph,

N_t. Number of time intervals in the study period,

Δt: Interval length in hr, and

TT_{SL}: Travel time through the control sections under the posted speed limits.

The second term in Equation 6-5 is equal to the total travel time under the posted speed limits. It is assumed that the minimum distance between the critical CMSs is 2 miles. The problem was submitted in NEOS server (Czyzyk et al. (1998), Dolan (2001)) and solved using LOQO 7.00, which is a solver for large-scale, nonconvex, and nonlinear optimizations.

6.2.1.1.1 Results for Greedy Algorithm

The greedy algorithm was implemented when CMSs are evaluated every 0.5 miles. In other words, CMSs are placed only in sections with even indices. Figure 6-2 displays how the locations of critical CMSs influence the objective function in Iteration 1. Sections are numbered in the opposite direction of traffic (See Figure 6-1). The most downstream section in the control zone is the section 2, and the most upstream section considered in the Greedy Algorithm is Section 100. The dashed line represents the objective function value correspond to Base condition when there is no CMS (n=0) and the objective function is equal to total travel time. Overall, as the CMS is placed closer to the bottleneck, the objective function and the total travel time become higher. Compared with Base condition, the objective function is not improved when the CMS is located at section 20 or downstream. This might be due to the fact that the distance between the CMS and queue are relatively short for these sections and there is not enough opportunity for speed harmonization system to influence traffic. For this iteration, CMS 100 is chosen since it returns the minimum objective function of $F_{\min}^{(1)}$=430.65 hr. Total delay values with and without speed harmonization are 27.59 veh-hr and 34.18 veh-hr, respectively and are computed using equation 6-5. This is corresponding to 19.28% $\left(=\frac{34.18-27.59}{34.18}\right)$ delay reduction.

For practical considerations, one may consider other locations in the neighborhood of the section 100 returning similar objective function values. For example in this case, sections 86 to 100 have very similar objective function values ranging from 430.24 veh-hr to 430.89 veh-hr. One may consider these sections in case practical considerations, such as restricted geometric condition, do not allow to place a critical CMS in a particular place.

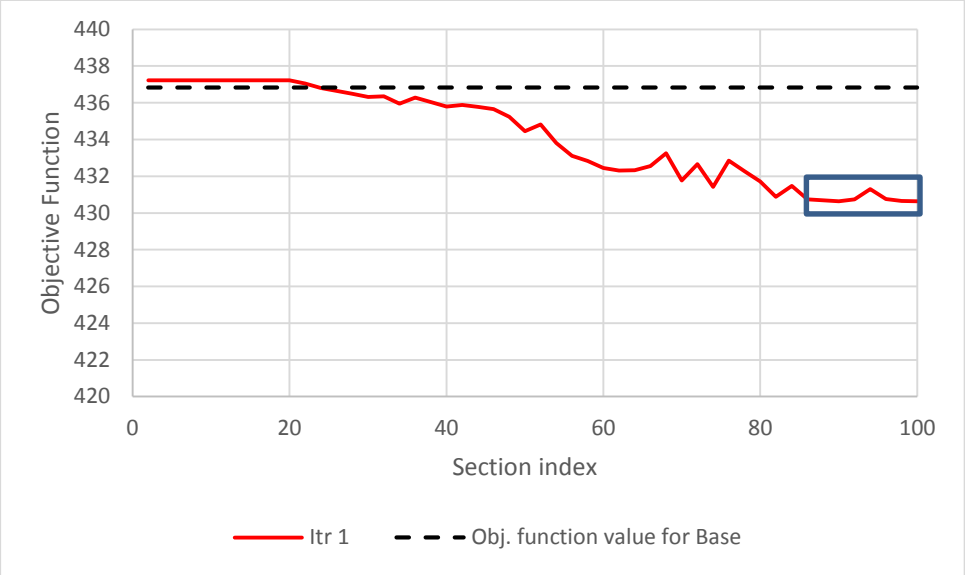


Figure 6-2: Greedy Algorithm results, iteration 1

In all the trials for Iteration 2, two critical signs are located in the control zone. One sign is always in Section 100 and the other sign is moved from one section to another such that minimum spacing between the signs is equal to or greater than 2 miles. This means that sections 2 to 92 are evaluated for the placement of the second CMS. Figure 6-3 shows variations of objective function when the location of the second critical CMS changes. Compared with Base condition, all the trials improved the objective function; however compared with Iteration 1, the objective function was not improved considerably for some of the locations. These sections are sections 60 to 92 and sections 2 to 22. This might be because the Sections 60 to 92 are far for placement of the second sign and as a result the dominant effect of speed reduction in these sections is increasing the total travel time rather than reducing shockwave propagation. For the section 2 to 22, the second sign does not have any impact because these sections are close to the congested sections and there is not enough opportunity to influence traffic. The maximum improvement happens when the second sign is placed at Section 48 which results in the objective function of $F_{min}^{(2)}$=427.85 hr and delay reduction of 28.6%($= \frac{34.18-24.4}{34.18}$). The stopping criterion does not hold because the objective function value improved in Iteration 2, (i.e. $F_{min}^{(2)}$=427.85 < $F_{min}^{(1)}$=430.65) and Iteration 3 starts.

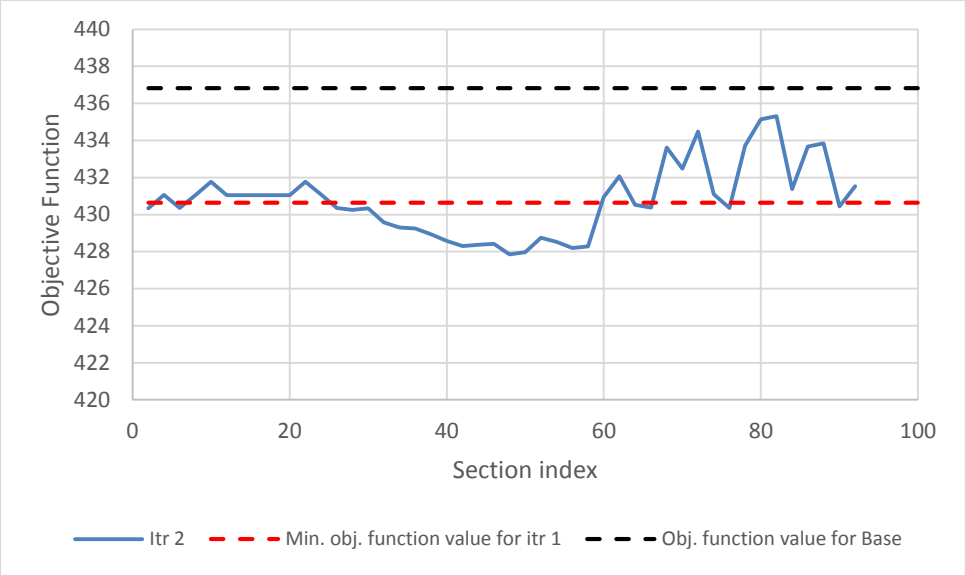


Figure 6-3: Greedy Algorithm results, iteration 2

In Iteration 3, two CMSs 100 and 48 are fixed and the third one is placed in the other sections with spacing of 2 miles or greater. Figure 6-4 shows variations of the objective function value when the location of the third sign changes. There is a discontinuity in Figure 6-4 because sections that are located less than 2 miles upstream and downstream of the Section 48 were not considered in the algorithm to satisfy minimum distance of 2 miles between two consecutive critical CMSs. Similar to Iteration 2, all the trials improve the objective function compared with Base condition; however there are few sections that improve the objective function compared with Iteration 2. The maximum improvement happens when the third sign is placed at section 40 returning objective function of $F_{min}^{(3)}$=427.5 hr which is corresponding to 30.8%($= \frac{34.18-23.64}{34.18}$) of delay reduction. The algorithm will proceed to Iteration 4 because $F_{min}^{(3)}$=427. 5< $F_{min}^{(2)}$=427.85

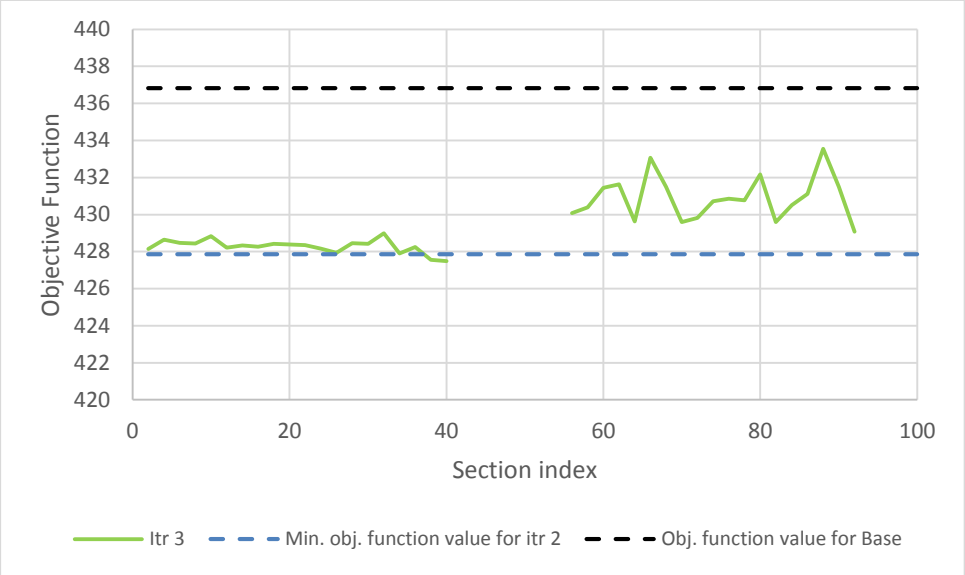


Figure 6-4: Greedy Algorithm results, iteration 3

All the trials in Iteration 4 reduce objective function compared with Base condition (See Figure 6-5); however they did not improve the objective function compared with the Iteration 3. The minimum objective function in iteration 4 is $F_{min}^{(4)}$ = 427.86 which is greater than $F_{min}^{(3)}$ =427. 5; thus the algorithm stops and returns sections 40, 48, and 100 as the optimal locations for CMSs.

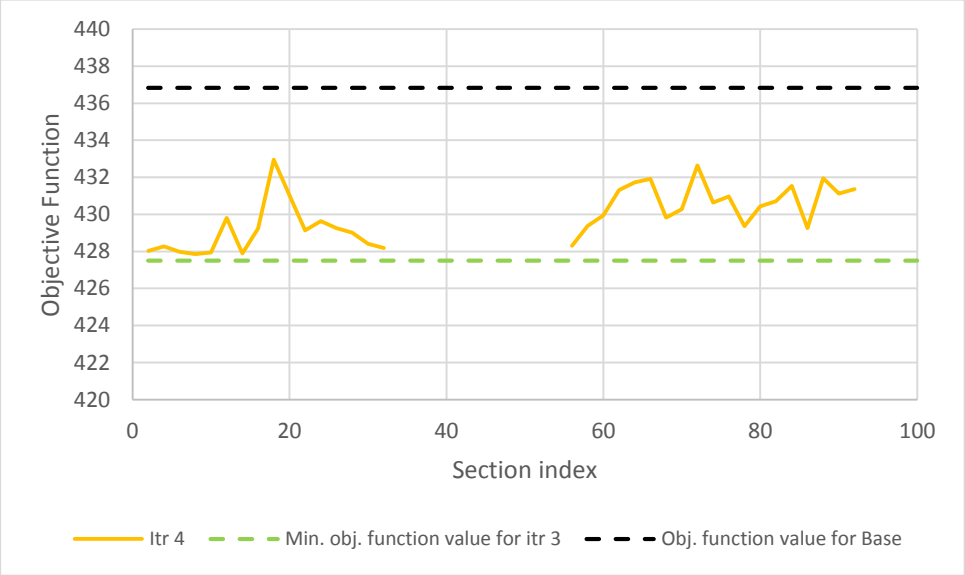


Figure 6-5: Greedy Algorithm results, iteration 4

6.2.2 Method 2: Augmented-Cut and Branch (AC&B)

The Branch and Bound Method is a well-known method to solve problems with integer variables and it has been further modified and improved by adding cutting planes. The cutting planes will be added in the form of extra constraints that restrict the feasible area to the portion that most likely includes the optimal and integer solution. A proper incorporation of a cutting plane, will cause fewer number of branching and will expedite solution process. If the cut is imposed in the root node of a tree the method is called Cut and Branch. Otherwise the method is called branch and cut.

The method 2 is an iterative algorithm, and in each iteration, the location of one critical sign is determined using Cut and Branch Method. The formulation of the cut changes from one iteration to another. For iteration n, the cut is formulated as

$$\sum_{i\in I} x_i = n \qquad 6\text{-}6$$

Where the variable x_i is the binary variable which is equal to 1 if a critical CMS is placed in section i; otherwise it is equal to zero. This cut forces the problem to have n critical signs in the solution. The n-1 of these critical signs have been already determined in the n-1 previous iterations. In the current iteration (i.e. Iteration n) the problem determines the location of n^{th} critical CMS. To determine the location of the next critical sign (i.e. $n+1^{st}$ critical sign), the cut is augmented by one unit and becomes

$$\sum_{i\in I} x_i = n + 1$$

Thus, the study called this method, Augmented-Cut and Branch (AC&B) since in each iteration the Cut and Branch Method is used and the cut is augmented by one unit from one iteration to another.

The overall steps of the method are shown in Figure 6-6. First, define P to be the formulated problem in Chapter 3. In Step 1, $P^{(n)}$ is constructed by adding the cut 6-6 to the formulation. In Step 2, the solution for $P^{(n)}$ is obtained using Branch and Bound method to determine the optimal location of the n^{th} critical CMS. In Step 3, the stopping criterion is checked. In this study, if the objective function increases or does not change, the algorithm will stop. If the stopping criterion is not met, Step 4 will place the n^{th} critical sign in the optimal location found in Step 3.Then, the algorithm starts Step 1 of Iteration n+1. Details about the steps of the method will be presented in sections 6.2.2.3. In the meantime, the next section will

discuss how branching rule and stopping criteria are adopted in the branch and cut algorithm (Step 2) considering the characteristics of the problem.

6.2.2.1 Branching rule

A heuristic approach is proposed for establishing the branching rule based on the concept of largest fractional value. Once the relaxed $P^{(n)}$ is solved, the values of the relaxed x_is are fractional and between zero and one. It is suggested to start branching from the x_i that has the largest value fractional value.

Rule 1: Start branching by selecting $\tilde{x}^{(n)}$ which is the variable x_i with the largest fractional value in iteration n.

This rule tries to pick the section that is more likely to have greatest influence to obtain the optimal objective function. Based on Constraint 5, the value of each x_i in the solution of the relaxed $P^{(n)}$ can be interpretted as the portion of speed change that the CMS i can implement to obtain the optimal objective function value. In particular, if x_i is zero, the CMS cannot change traffic speed; and if it is one, the CMS has maximum influence by implementing maximum speed change. Thus, a section with the maximum fractional value can be interpreted as the location which has the greatest influence on the objective function.

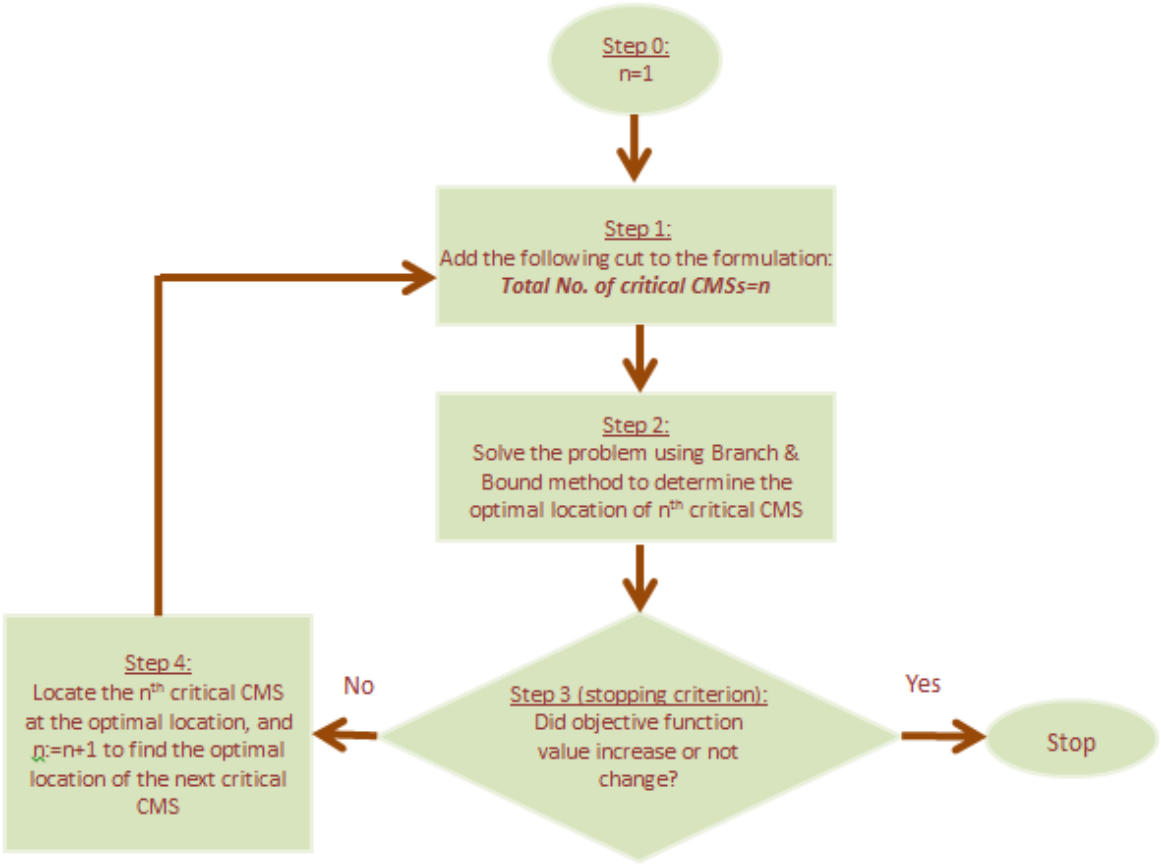


Figure 6-6: Steps of AC&B method

Once the branching starts there might be several variables to select and continue branching. The Rule 2 indicates how to continue branching.

Rule 2: If $\tilde{x}^{(n)}$ is already evaluated, select another x_i that has not been evaluated yet and

a) is the closest (in terms of the distance) to the incumbent section , and

b) is within a user-defined window (Fixed Window Strategy) . The suggested window has m-1 sections downstream and m-1 sections upstream of the incumbent section where m is the minimum number of sections between two consecutive critical CMS. .

The Rule 2 is used to ensure that the critical CMS location, chosen in Iteration n, is better than the neighboring sections. The neighboring sections are selected to be m-1 sections upstream and m-1 sections downstream of the critical CMS because these neighboring sections will be eliminated in the next iteration to preserve the minimum distance between the consecutive

critical signs. This way, the algorithm ensures that the eliminated sections are not better than the critical CMS.

The iteration will stopped when the sopping strategy discussed in the following section is met.

6.2.2.2 Stopping strategy

A heuristic procedure, called Fixed Window Strategy, is proposed to stop each iteration and then proceed to the next iteration or stop the algorithm.

When an incumbent solution is the best solution compared with w sections upstream and w sections downstream, then the branch and bound algorithm stops in an iteration.

When w is equal to 1, the Fixed Window Strategy is translated to the stopping criteria for a strictly convex problem where the global optimal solution is better than the immediate neighbors. . Still this strategy can be used to find a local optimal solution for a non-convex problem, but this may lead to a local optimal solution that results in insignificant improvement. Thus w should be more than 1 to reduce the likelihood of trapping in impaired local optimal solutions. The study suggests w to be m-1 where m is the minimum number of sections between two consecutive signs. This is decided with consideration of Constraint 10 (i.e. minimum spacing constraint) that states no critical CMSs should be placed within the upstream and downstream sections of a Critical CMS. For optimal performance this implies that the location of a critical CMS is better than m-1 sections upstream and m-1 section downstream

6.2.2.3 Details of the AC&B Method

Let $P_r^{(n)}$ be the relaxed version of the $P^{(n)}$ where all binary variables are converted to continuous variables between 0 and 1. Also assume F_U be the upper bounds for the objective function and TTT_{Base} be the total travel time in the Base condition. The Steps presented in Figure 6-6 are expanded as below:

Step 0 : Set up $F_U^{(0)} = TTT_{Base}$ and n=1

Step 1: Construct $P^{(n)}$ by adding the following cut to P:

$\sum_{i \in I} x_i = n$

Step 2: Solve $P^{(n)}$ using branch and bound method

Step 2.1: Solve $P_r^{(n)}$ and $F_U^{(n)} = F_U^{(n-1)}$

Step 2.2: Branch: Based on a branching rule, select a node and solve $P^{(n)}$. Section 6.2.2.1 proposed a branching rule.

Step 2.3: Tighten the upper bound: Find the incumbent solution which is the solution returning the minimum objective function. Then update the upper bound as:

$$F_U^{(n)} = \min\{F_U^{(n)}, \text{objective function of the incumbent solution}\}$$

Step 2.4: Fathoming: Nodes should be excluded from further branching

1) Its objective function value is more than the F_U
2) The node is not feasible
3) The solution for $P_r^{(n)}$ has integer x_i's.

Step 2.5: Stop the procedure if

1) There is no node to evaluate
2) Fixed Window Strategy is satisfied.

Step 3: if $F_U^{(n-1)} \leq F_U^{(n)}$, stop; otherwise fix the values of x_is that are equal to 1; and n=n+1, and go to Step 1.

6.2.2.4 Benchmark problem results for AC&B Method

The benchmark solution introduced in section 6.2.1.1 is solved using AC&B Method. According to the algorithm, the $P^{(1)}$ is constructed by adding the cut $\sum_{i\in I} x_i = 1$ to the program formulated in Chapter 3. The relaxed version of the program, $P_r^{(1)}$, is solved. The resulting fractional values of x_i's are scaled up by factor of 100 and shown in Figure 6-7. The value of x_{101} in Figure 6-7 is the highest; thus this section is selected to start branching. Since the section 101 is the most upstream section, the branching is possible only for the downstream sections. The Figure 6-8 shows the results of feasible branches for different iterations. The gray cell represents the section with the maximum value among continuous x_i's; thus the cell corresponding to $x_{101}$is gray. Objective function values decrease as the algorithm moves from section 101 to section 99. Thus x_{99}=1 is an incumbent solution. Branching for 7 (=m-1) sections downstream of the incumbent solution will not reduce the objective function. Hence the Fixed Window Strategy is satisfied and the algorithm stops branching. Section 99 is selected for the location of the first critical CMS resulting objective function of 431.29 and delay reduction of

17.4% compared to Base condition (i.e. when there is no speed harmonization). The algorithm will proceed to the next iteration since $F_U^{(1)} = 431.29 < F_U^{(0)} = 436.83$.

For the second iteration, x_{99} is fixed to be 1; x_i's for Sections 92 to 98 as well as Sections 100 and 101 are fixed to be zero to preserve the minimum distance of 2 miles for critical CMSS; and the other x_i's are relaxed. The cut is augmented to determine the location of the second sign. The relaxed problem $P_r^{(2)}$ is solved and fractional values of x_i's (see Figure 6-7) indicated that the variable x_{59} has the maximum value and branching starts from this variable. Branching for the 7 sections upstream of Section 59 deteriorates the objective function and is not further continued. Branching for the downstream sections returns better solutions until Fixed Window Strategy is satisfied when x_{50}=1. The corresponding objective function is 428.89, resulting 25.6% of delay reduction. The algorithm does not stop because $F_U^{(2)} = 428.89 < F_U^{(1)} = 431.29$.

In the third iteration, x_{21} has the maximum fractional value as shown in Figure 6-7 . Branching for the 7 downstream sections will increase the objective function, thus the algorithm does not further branch for the downstream sections. The minimum objective function value of 427.42 is achieved when x_{38}=1. The corresponding section is selected for the location of the third sign and with 31.1% of delay reduction. In this iteration, the method does not evaluate sections 43 and upstream to maintain the minimum distance of 2 miles from CMS 50. The algorithm will proceed to Iteration 4 because $F_U^{(3)} = 427.42 < F_U^{(2)} = 428.89$.

In the fourth iteration, x_{21} has again the maximum fractional value (see Figure 6-7) and branching starts from this section and will be continued for upstream and downstream sections. The minimum objective function value is achieved when $x_{28} = 1$, and it is equal to 428.08. The algorithm stops because $F_U^{(4)} = 428.08 < F_U^{(3)} = 427.42$. This means that installing the fourth sign will not reduce the total travel time by the marginal penalty of 0.403 or more. The algorithm does not evaluate sections 31 and upstream to maintain the minimum distance of 2 miles from CMS 38.

Thus the AC&B method determined 3 signs to be installed in sections 99, 50, and 38 resulting in 31.1% delay reduction.

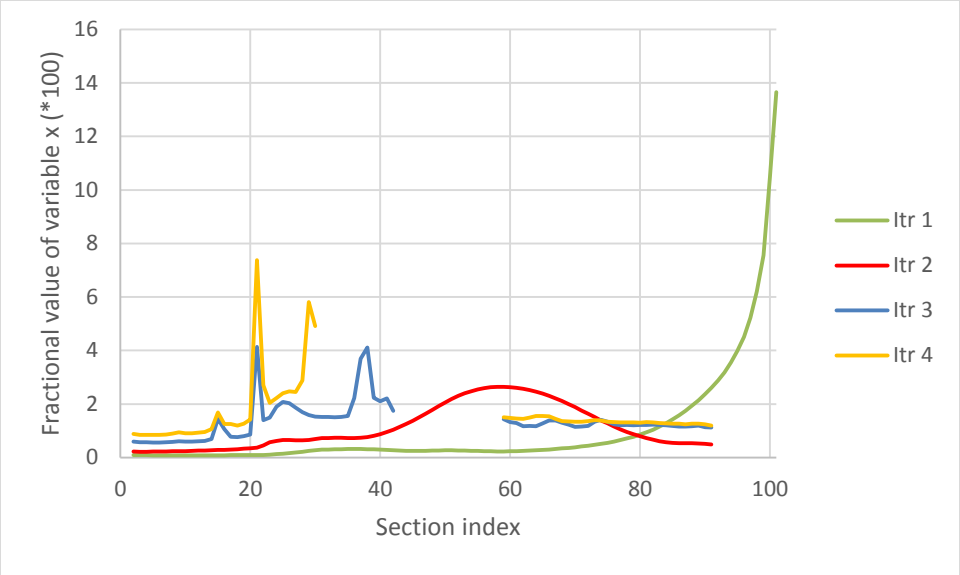


Figure 6-7: Values of continuous x_is in different iterations

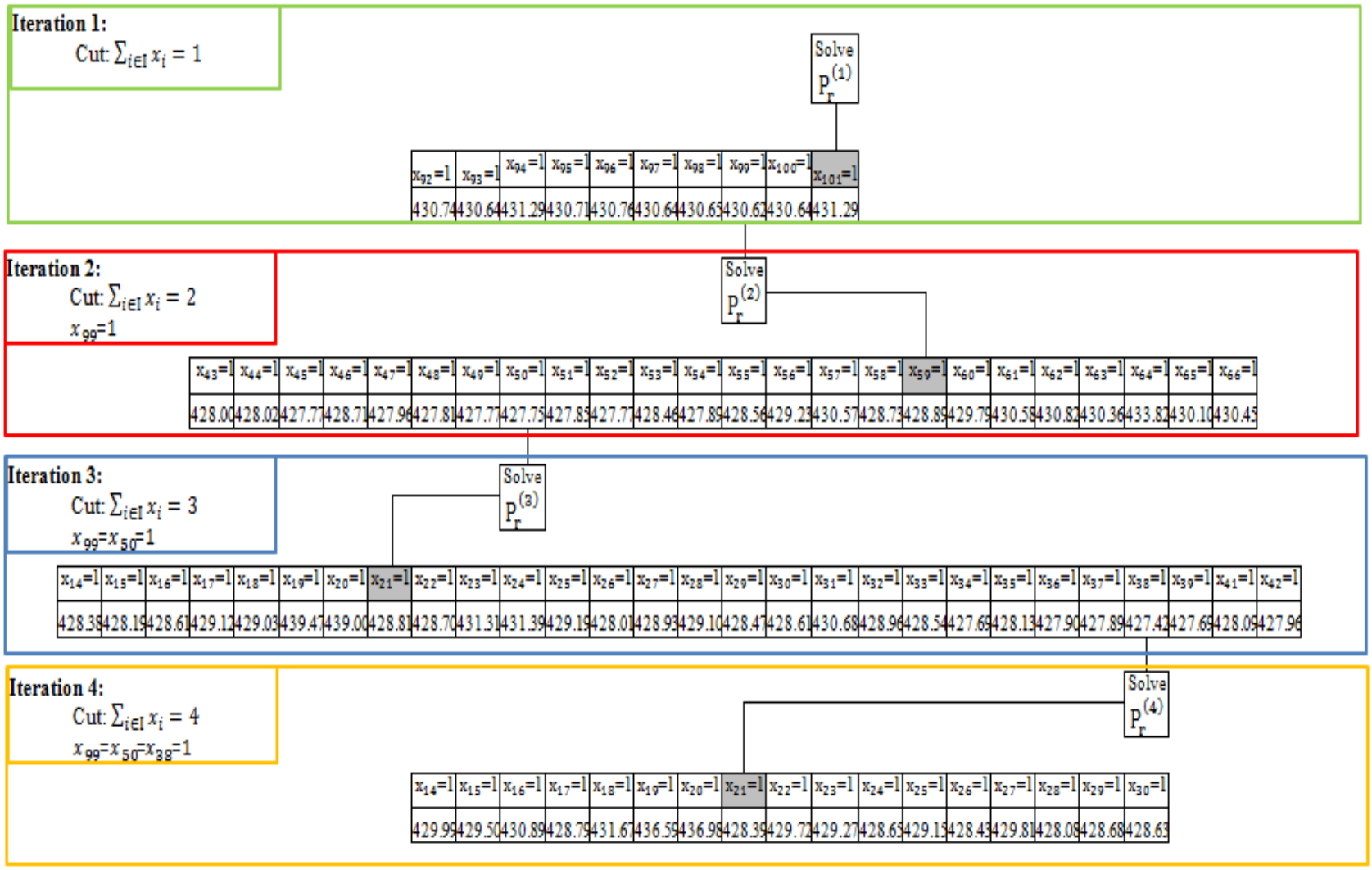


Figure 6-8: Results of AC&B method for the benchmark problem

6.2.3 Method 3: Approximate Decomposition

Method 2, Augmented Cut and Branch, sometimes require many branching and consequently significant computational resources. For instance, the benchmark problem was solved after solving 80 feasible branches which might be computationally expensive. This sets the motivation to develop a solution method which requires less computational resources. The Approximate Decomposition method is proposed to be computationally more efficient than the AC&B method. In Approximate Decomposition method the problem is decomposed into two sub-problems that are sequentially solved:

Sub-problem 1: Speed Harmonization using Mini CMSs

Sub-problem 2: Mini CMS Packing Problem

The first sub-problem is a nonlinear program which determines the optimal advisory speeds when all sections have a sign, called Mini Changeable Message Signs (MCMSs). MCMSs have less speed reduction capability than a critical CMS. The next section will formulate the relationship between maximum speed reduction capabilities of Mini CMSs and critical CMSs. The second sub-problem packs mini CMSs and replaces them by a critical CMS such that it yields similar speed reduction effects. More details on sub-problems are explained in subsequent sections.

6.2.3.1 Sub-Problem 1: Speed Harmonization using Mini CMSs

In this sub-problem, one MCMSs is located at upstream boundary of each section. Assume, maximum speed reduction that can be implemented by a MCMS is denoted by Δ'_s, which is less than maximum speed reduction of Δ_s implemented by critical CMSs. This section intends to determine the relationship between Δ'_s and Δ_s such that final solution satisfies the integral constraints 10 which determine the minimum spacing between the critical signs. Then, one can eliminate the integral constraint and solve the problem in a continuous space.

6.2.3.1.1 Relationship between Δ'_s and Δ_s

Figure 6-9 shows a roadway with critical CMS (part a) and with MCMS (part b) where traffic moves from right to left and sections are numbered from downstream to upstream. In Figure 6-9a, two critical CMSs are installed in Sections i and i-k and in Figure 6-9b, all the sections have a MCMS. It is intended to establish a relationship between Δ_s and Δ'_s such that the two CMS types would yield similar travel speeds between Section i and Section i-k. To do so, a

condition is considered that both systems display the maximum speed change. The vehicle with speed of V enters Section i and the CMS i reduces the speed of the vehicle by Δ_s; thus the travel speed between the two critical CMSs i and i-k is $V - \Delta_s$. When mini CMSs are installed, each MCMS reduces speed by Δ'_s. The travel speed for the section i is $V - \Delta'_s$ and this travel time is further reduced by Δ'_s per section. Average speed between the two sections i and i-k can be approximated by the arithmetic mean of the travel speeds of in-between sections and will be equal to $V - \frac{1+k}{2}\Delta'_s$. The average speed when MCMSs are installed should be equal to the average speed when critical CMSs are installed, so

$$V - \frac{1+k}{2}\Delta'_s = V - \Delta_s \qquad 6\text{-}7$$

With algebraic manipulation:

$$\Delta'_s = \frac{2}{1+k}\Delta_s \qquad 6\text{-}8$$

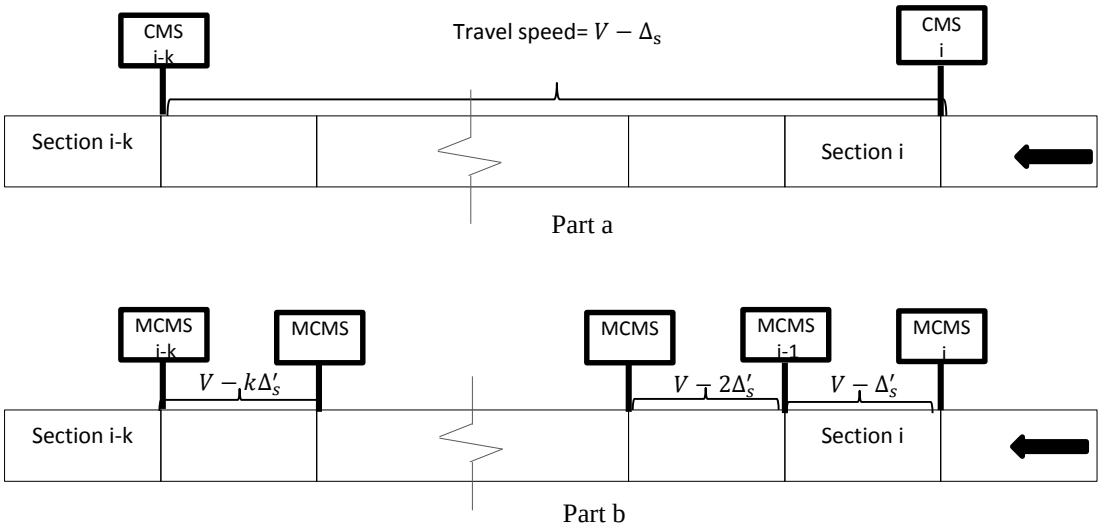


Figure 6-9: A Roadway with a) Critical Signs b) Mini Signs

In equation 6-8, value of Δ'_s increases as k decreases. The upper bound for Δ'_s is achieved when k is equal to the minimum number of sections between two consecutive signs denoted by m. Thus the upper bound for Δ'_s is:

$$\Delta'_s \leq \frac{2}{1+m}\Delta_s \qquad 6\text{-}9$$

If a value greater than $\frac{2}{1+m}\Delta_s$ is chosen for Δ'_s, the packing process in the second sub-problem may violate the minimum allowable distance between the two critical signs. Smaller values of Δ'_s would create smoother advisory speed trajectories and would help reducing the number of signs. Thus the study recommends determining the number of signs and their locations based on a small value for Δ'_s (e.g. 0.25 mph) and then incrementally increasing its value within the range indicated in relationship 6-9; one can stop the algorithm when increasing Δ'_s does not improve the objective function. The effect of Δ'_s on the number and spacing between the signs will be explored by solving a benchmark problem in section 6.2.3.3.

Assuming a Δ'_s is selected within the range indicated in relationship 6-9, one can solve the Sub-problem 1 by the following modifications:

- Modify the original objective function by removing the penalty term for the number of signs because in the sub-problem 1 the number of signs is known (i.e. every section has a sign) and the advisory speeds are the only decision variables.

$$\text{Min}((\sum_i \sum_t D_{i,t} * \lambda_i * L * \Delta t) \qquad 6\text{-}10$$

- Eliminate Constraint 10
- Modify Constraint 5 as

$$-\text{MaxAS}_i \le (AS_{i+1,t} - AS_{i,t}) \le \Delta'_s \qquad \forall i \in I, \forall t \in T \qquad 6\text{-}11$$

Compared with the original Constraint 5, we removed integer variables and changed the upper bound from Δ_s to Δ'_s.

The advisory speeds determined by the sub-problem 1 will be used to solve the sub-problem 2.

6.2.3.2 Sub-Problem 2: Mini CMS Packing Problem

The purpose of this sub-problem is to pack the mini CMSs and replace each pack by only one critical CMS that is located at the upstream boundary of packed sections yielding similar travel speed (influence) on traffic. For each mini CMS i, the study defines an influence index, $I^m_{i,t}$, which changes over time as

$$I^m_{i,t} = AS^m_{i+1,t} - AS^m_{i,t} \le \Delta'_s \qquad 6\text{-}12$$

Where, $AS^m_{i,t}$ is the advisory speed displayed by MCMS i during time interval t.

$I_{i,t}^{m}$ is associated with the speed drop implemented by the MCMS i. If the advisory speed displayed be a MCMS is equal to the one displayed by the upstream MCMS, the MCMS does not have any influence on traffic and $I_{i,t}^{m}$ will be equal to zero. If the influence index is positive, the MCMS reduces travel speed; and if it is negative, traffic may increase speed.

To derive a packing criteria, Figure 6-10 shows a network with MCMSs that can have different influence indexes. The travel speed for a section l is computed $V - \sum_{j=l}^{j=i} I_{j,t}^{m}$. Consequently, the average travel speed can be approximated as arithmetic average of the travel speeds in each section as

$$\overline{V} = \frac{\sum_{l=i-k+1}^{l=i}(V-\sum_{j=l}^{j=i} I_{j,t}^{m})}{k} \quad 6\text{-}13$$

This can be further simplified:

$$\overline{V} = V - \frac{\sum_{j=i-k+1}^{j=i}(j-i+k)I_{j,t}^{m}}{k} \quad 6\text{-}14$$

Similarly, for the critical CMS i, the influence index is defined as

$$I_{i,t} = AS_{i,t} - AS_{i-k,t} \leq \Delta_s \quad 6\text{-}15$$

Where, i and i+k are the section indices for two consecutive critical signs.

This defines the influence index as the difference between advisory speeds displayed by two consecutive critical CMSs.

The travel speed between the two consecutive critical CMSs is

$$\overline{V} = v - I_{i,t} \quad 6\text{-}16$$

The condition to pack k mini CMSs is that the amount of influence (i.e. speed reduction) produced by MCMSs should be equal to or less than the amount of influence (i.e. speed reduction) produced by critical CMSs. Thus the following condition should be satisfied:

$$\frac{\sum_{j=i-k+1}^{j=i}(j-i+k)I_{j,t}^{m}}{k} \leq I_{i,t} = \Delta_s \quad 6\text{-}17$$

The relationship 6-17 means that the weighted average of influence indexes of MCMSs should be less than or equal to maximum influence index of critical CMSs. The upstream section in the pack has the weight of k and the most downstream section in the pack has the weight of 1. This is intuitive because any speed reduction in the upstream section i will be traversed to downstream sections and will persist longer; thus its weight is equal to number of sections receiving that influence and should be larger in amount.

Since this relationship has to hold for all time intervals then

$$\max_{t}\left(\frac{\sum_{j=i-k+1}^{j=i}(j-i+k)I_{j,t}^{m}}{k}\right) \leq \Delta_s \qquad 6\text{-}18$$

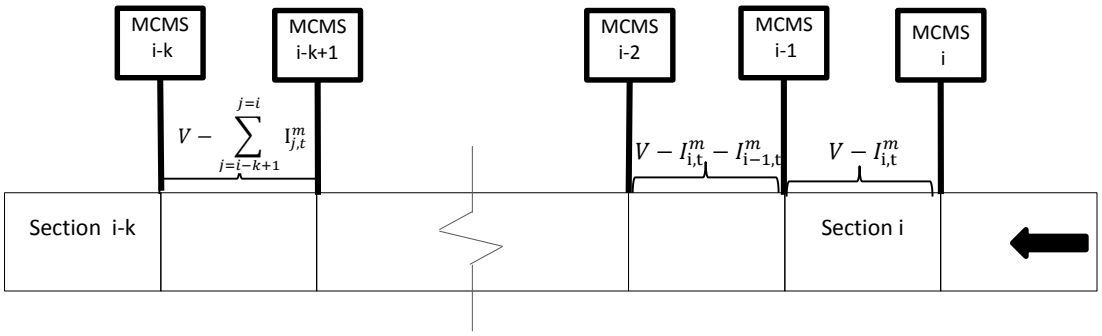


Figure 6-10: Influence indexes for Mini CMSs for a general condition

Two heuristic algorithms are presented to solve the mini CMS packing problem

6.2.3.2.1 Heuristic algorithm 1: forward packing

This method starts packing from the most upstream section N. We put a sign at the upstream boundary of section N, and keep adding sections to the corresponding pack until $\max_{t}\left(\frac{\sum_{j=k}^{j=N}(j-k+1)I_{j,t}^{m}}{N-k+1}\right)$ becomes greater than Δ_s. At this moment, another critical sign is placed at upstream section of k and a new pack is created. The first section in the new pack will be Section k and consecutive sections will be added until the total influence index exceeds Δ_s and this process will continue till all the mini CMSs are packed. The algorithm is presented as below:

Define n_i to be the section index for the i^{th} critical sign (i is a counter for critical signs)

Step 0: Put a sign at upstream boundary of section N, and let $n_1 = N$, i=1, and k=n_1

Step 1: k := k-1 and compute $\max_{t}\left(\frac{\sum_{j=k}^{j=n_i}(j-k+1)I_{j,t}^{m}}{n_i-k+1}\right)$

Step 2: If

$$\max_{t}\left(\frac{\sum_{j=k}^{j=n_i}(j-k+1)I_{j,t}^{m}}{n_i-k+1}\right) \leq \Delta_s$$

Then k:=k-1; otherwise i=i+1 and $n_i = k$

Step 3: If k≠1 go to Step 1; otherwise stop.

6.2.3.2.2 Heuristic algorithm 2: backward packing

The logic for the backward packing is the same as the logic for forward packing except it starts from very downstream section and moves in opposite direction of traffic. The algorithm is presented below:

Define n_i to be the section index for the i^{th} critical sign, and let b_i be the section index for the most downstream section in pack i.

Step 0: k=1, i=1, $b_i = 1$

Step 1: If k<=N then compute $\max_t\left(\frac{\sum_{j=b_i}^{j=k}(j-b_i+1)I_{j,t}^m}{k-b_i+1}\right)$

Step 2: If k=N+1 or $\max_t\left(\frac{\sum_{j=b_i}^{j=k}(j-b_i+1)I_{j,t}^m}{k-b_i+1}\right) > \Delta_s$ then $n_i = k-1$, i=i+1, and $b_i = k$

Step 3: If k<=N then k:=k+1 and go to Step 1; otherwise stop.

6.2.3.3 Benchmark problem solution

The benchmark problem introduced in Section 6.2.1.1 is solved with different values of Δ_s' to explore its effect on the optimal solution. In general, one may select a Δ_s' from the following range:

$$0.25 \leq \Delta_s' \leq \frac{2}{1+m}\Delta_s \qquad 6\text{-}19$$

The upper bound is established based on Eq. 6-9. The lower bound of 0.25 mph is selected because a Δ_s' of less than 0.25 may not create the speed reduction that is required for speed harmonization.

The effect of Δ_s' will be evaluated for the benchmark problem when $\Delta_s = 5$ and $m = 8$. Thus the range for Δ_s' will be

$$0.25 \leq \Delta_s' \leq 1.11 \qquad 6\text{-}20$$

The problem is solved for Δ_s' of 0.25, 0.5, 0.75, and 1. For each Δ_s', two different solutions based on Forward Packing and Backward Packing are derived. The signs locations, packs and corresponding influence indexes are reported in Table 6-3. For instance, when $\Delta_s' = 1$, Forward Packing returns 6 signs in sections 101, 90, 67, 37, 27, and 18. The influence indexes for packs 1 to 5 are close to 5 and there is only pack 6 that has a very low influence index because this pack was created at the end of the forward packing algorithm and includes the residual effects of mini signs.

According to results in Table 6-3, the value of Δ_s' has correlations with the number of signs. As Δ_s' reduces, the number of signs either stays the same or reduces. This is because of two reasons:

- First, Δ_s' determines maximum number of critical signs and reducing Δ_s' could reduce the number of the critical signs. The maximum number of critical signs is needed when all mini CMSs are active and display maximum speed reduction of Δ_s'. The following relationship holds between maximum number of signs, ñ, and minimum number of sections, k_{min}, between two consecutive signs.

  $$\tilde{n} = \left\lceil \frac{L_{control}}{k_{min}} \right\rceil \quad 6\text{-}21$$

  Where, $L_{control}$ is the length of the control zone over which critical CMSs can be potentially placed.

  According to Equation 6-8, the minimum number of sections, k_{min}, between two consecutive critical CMSs is computed as.

  $$k_{min} = \left\lceil \frac{2\Delta_s}{\Delta_s'} - 1 \right\rceil \quad 6\text{-}22$$

One can deduce from Equations 6-21 and 6-22 that the number of critical signs reduces as Δ_s' is reduced.

- The second reason is related to the effect of Δ_s' on the quality of packing. When Δ_s' is smaller, the packing algorithms are more efficient. In theory, the maximum unused space in the pack is $\Delta_s' - \delta$ where, δ is a very small number. For smaller Δ_s', packing is expected to be more efficient because maximum unused space becomes smaller; moreover, influence indexes are smaller and more sections can be placed in a given pack. This means that the number of required packs (critical signs) is expected to be lower since the number of sections is fixed.

  The results in Table 6-3 show that the minimum objective function was obtained when Δ_s' = 0.25 and signs are determined by forward packing.

6.2.3.3.1 Recommendations for choosing Δ_s' and packing algorithms

When there are limitations for computational resources such that the parallel problem solution is not possible, the study recommends solving the problem using a small value for Δ_s'

(e.g. 0.25 mph); and then incrementally increasing its value within the range indicated in relationship 6-9. One can stop the algorithm when increasing Δ'_s does not improve the objective function. This approach requires the problem to be solved sequentially for different values of Δ'_s and as a result may elongate solution time.

When computational resources are available, one can run the problem in parallel for different values of Δ'_s and choose the solution corresponding to best objective function. This requires exhaustive search within the range indicated in relationship 6-9. This is not expected to be a large range as for the benchmark problem, Δ'_s ranges from 0.25 mph to 1.11 mph.
Forward packing is compared to backward packing as follow:

1) Objective function value comparison: For Δ'_s of 0.5 and 1, the backward packing algorithm returned a lower objective function value; while for Δ'_s of 0.25 and 0.75, the forward packing returned a lower objective value. The difference between the objective function values (forward-backward) ranges from -1.02 veh-hr to 4.01 veh-hr corresponding to -0.23% to 0.92% of total travel time (436.83 veh-hr) when there is no speed harmonization. Thus, none of the methods consistently returns a better objective function value and the difference between objective function values is less than 1%.
2) Number of packs (signs): For Δ'_s of 0.75 and 1, the forward packing computes one more sign than the backward packing, but for Δ'_s of 0.25 and 0.5, the two methods return the same number of signs. For small values of Δ'_s, the two packing methods return similar number of packs possibly because the size of items are smaller leading to less variations in the used space of each pack.
3) Incomplete packs: Incomplete packs are defined as the packs that their unused space is equal to 1 mph or higher. This means that the incomplete packs have a used space (i.e. influence index in Table 6-3) of 4 or lower. The incomplete pack may be created at the end of packing algorithm since both forward and backward algorithms try to completely fill a pack before using a new pack. The backward packing method returns only one incomplete pack for Δ'_s of 0.5 with used space of 2.19 mph. The forward packing method creates three incomplete packs for Δ'_s of 0.5, 0.75 and 1. The corresponding used space ranges from 0.13 mph to 2.64 mph.

   For Δ'_s of 0.75 and 1, the forward packing returned 1 more signs than the backward packing and these signs are corresponding to the incomplete packs. Overall, one sign

difference may not be practically the major factor to differentiate the two methods especially for this benchmark problem in which the two methods yield very close objective function values.

Based on the comparisons above, the backward packing and forward packing methods retuned very close objective function values, but for two cases out of four, the backward packing algorithm returned one less sign than the forward packing algorithm. Thus in the case of limited computational resources, one may use backward packing algorithm; however if there is not such a limitation and there is a possibility for parallel solution, it is recommended using the two algorithms and find the solution with the lowest objective function. This conservative recommendation will increase the possibility of finding a better solution.

Table 6-3: Results of Approximate Decomposition method

	Forward Packing				**Backward Packing**			
Δ_s'	**Sign* location**	**Pack****	**Pack influence index**	**Objective*** function**	**Sign* location**	**Pack****	**Pack influence index**	**Objective*** function**
1	101, 90, 67, 37, 27, 18	P1: [91,101] P2: [68,90] P3: [38,67] P4: [28,37] P5: [19,27] P6: [2,18]	P1: 4.76 P2: 4.96 P3: 4.94 P4: 4.82 P5: 4.86 P6: 0.4	429.67	25, 35, 54, 87, 99	P1: [2,25] P2: [26,35] P3: [36,54] P4: [55,87] P5: [88,99]	P1: 4.91 P2: 4.97 P3: 4.94 P4: 4.85 P5: 4.76	429.30
0.75	101, 86, 58, 38, 23	P1: [87,101] P2: [59,86] P3: [39,58] P4: [24,38] P5: [2,23]	P1: 4.75 P2: 4.95 P3: 4.80 P4: 4.88 P5: 2.64	428.30	27, 43, 80, 97	P1: [2,27] P2: [28,43] P3: [44,80] P4: [81,97]	P1: 4.93 P2: 4.95 P3: 4.88 P4: 4.73	429.32
0.5	101, 77, 35, 12	P1: [78,101] P2: [36,77] P3: [13,35] P4: [2,12]	P1: 4.90 P2: 4.96 P3: 4.95 P4: 0.13	433.07	31, 53, 86, 101	P1: [2,31] P2: [32,53] P3: [54,86] P4:[87,101]	P1: 4.72 P2: 4.83 P3: 4.91 P4: 2.19	429.06
0.25	101, 53	P1: [54,101] P2: [2,53]	P1: 4.93 P2: 4.93	427.94	52, 100	P1: [2,52] P2: [53,101]	P1: 4.89 P2: 4.90	428.75

*They are written in the order that the algorithm determines.

**Packing algorithm applied for sections 2 to 101 that are either in the work zone or upstream (see Figure 6-1)

*** Includes penalty for the number of signs

6.3 Methods comparison and recommendations

The methods are compared in terms of 1) results including objective function, corresponding optimal solution, and percentage of delay reduction, and 2) number of runs which is an indication of requires computational resources and time to solve the problem.

6.3.1 Objective function values and delay

This chapter presented three methods to solve the sign location problem: 1) Greedy Algorithm, 2) AC&B Method, and 3) Approximate Decomposition Method. Table 6-4 compares the three methods in terms of the objective function, optimal solution, and resulting percent delay reduction.

The Greedy Algorithm and AC&B method returned similar results. The discrepancies in objective function value and delay reduction are roughly 0.08 veh-hr, and 0.25%, respectively. The locations of the critical signs are different by at most two sections (0.5 miles).

The results of the Approximate Decomposition is a little less optimal than the results of AC&B method. The discrepancy between the objective function value and percentage of delay reduction is 0.52 veh-hr and 2.7%, respectively. The number of required critical CMSs returned by AC&B Method is more than the Approximate Decomposition Method.

Table 6-4: Results of Approximate Decomposition method

Method	Optimal objective function value	% delay reduction	Optimal Locations (section indexes)	Number of problem solutions				
				itr 1	itr 2	itr 3	itr 4	Total
Greedy Algorithm	427.50	30.83	40,48,100	50	46	39	35	170
AC&B	427.42	31.05	38,50,99	10	24	29	17	80
Approximate Decomposition	427.94	28.34	53,101	_	_	_	_	12

6.3.2 Number of runs

The greedy algorithm requires the highest number of runs and the other methods reduced the number of runs and as a result, less computational resources are needed. The number of runs for AC&B method reduced by roughly a factor of 2.1 $\left(\approx \frac{170}{80}\right)$ and the number of runs for Approximate Decomposition method reduced by a factor of 14.1 $\left(\approx \frac{170}{12}\right)$.

The effects of following parameters are discussed on the number of runs.

- Number of sections (N_s): In greedy algorithm, the number of runs directly depends on the number of sections. For instance, for a network with 200 sections, the first iteration required 100 runs because the accuracy for sign placement was 2 sections (i.e. 0.5 mile). In AC&B method, the number of sections determines the upper bound for the number of runs. In general, the number of runs for the AC&B method is not equal to the upper

bound. For example, in the benchmark problem the number of runs for the first iteration is 10 while N_s is equal to 100. In the Approximate Decomposition method, the number of sections (N_s) does not have any influence on the number of runs.

- Optimal number of critical signs (n): both greedy algorithm and AC&B method, are iterative and the number of iterations is equal to n+1. The higher number of critical signs requires higher number of iterations and more runs. The Approximate Decomposition method is not sensitive to n. This parameter, in conjunction with number of sections determine an upper bound for the number of required runs in Greedy algorithm and AC&B method, and consequently determine the required computational resources. Number of optimal signs multiplied by the number of sections establishes an upper bound for the number of runs; however this bound is not tight, especially for the AC&B method. This may happen because some of the sections may be eliminated from evaluation to preserve the minimum spacing between the critical CMSs or the AC&B method finds a local optimal and stops.
- Maximum spatial change in advisory speed (Δ_s) : this does not have any influence on the number of runs for the Greedy Algorithm and AC&B method, but according to relationship 6-8 it linearly influences the upper bound of Δ_s'. For the benchmark problem, Δ_s was selected to be 5 mph and resulted in 12 runs. This includes four runs to solve the first sub-problem, four runs to determine optimal objective function values for the forward packing method, and similarly four runs for the backward packing algorithm. If someone selects the Δ_s of 10 mph, the maximum number of runs will be 24.
- The search increment for Δ_s': a small search increment will increase the number of runs for Approximate Decomposition method. The search increment for the range indicated in relationship 6-13 was 0.25, which results in 12 runs. If the search increment is 0.1, the number of required runs will be 33.

In the next section recommendations are provided to select the solution methods

6.3.3 Type of objective function

The approximate decomposition method was developed for travel time minimization. In particular, the criterion to replace a pack of mini CMSs with a Critical CMS is that the Critical

CMS is able to create similar travel time values as the corresponding pack creates. This criterion is based on travel time because the objective function is to minimize total travel time. For AC&B method and Greedy Algorithm, there is no assumption about the type of objective function. Thus these methods may be applied for other types of objective functions.

6.3.4 Recommendations to use the methods

Based on the discussion provided for the comparison of the methods, it is recommended the Approximation Decomposition method be used for travel time or delay minimization since it requires less number of runs. For other objective functions, the AC&B method can be used since it requires less number of runs than the Greedy Algorithm.

Chapter 7 Conclusions and recommendations

This chapter first presents the conclusions for the second order traffic stream model followed by the conclusions for the Work Zone Speed Harmonizer (WZSH). WZSH is a mathematical program that simultaneously determines optimal advisory speeds, location of critical Changeable Message Signs (CMS), and number of critical CMSs for speed harmonization. Then, the recommendations of the study are presented.

7.1 Second order traffic stream model

To the best knowledge of the author, this is the first study that calibrated relaxation time (τ) and anticipation coefficient (ϑ) for work zone traffic conditions. The parameters were determined to match speed and queue length data collected from the one-lane section of a freeway work zone. A new calibration methodology was proposed in two stages:

Stage 1: The methodology detected the behavior of the second order model in the τ - ϑ space and accordingly excluded the areas that did not include appropriate parameter values. It was found that the ratio of $\frac{\tau}{\vartheta}$ has the main influence on speed discrepancy and results showed, for $\frac{\tau}{\vartheta} \geq 10$ (τ in sec and, ϑ in $\frac{\text{miles}^2}{\text{hr}}$) the model overestimates speed and for $\frac{\tau}{\vartheta} \leq 4$, the model underestimates speed. In addition it was shown that beyond the threshold of 13 seconds, the relaxation time does not have a significant influence on speed estimates when section length is 450 ft. Smaller thresholds were found for shorter section lengths. Thus these areas were excluded from the search in Stage 2. As a result, the search area was limited to that portion of the τ - ϑ space that is likely to return acceptable discrepancy.

Stage 2: A local search inside the area narrowed down in stage 1, was conducted to find parameter values that return acceptable estimates of queue length and speed. The lesson learned from this stage is that the search area should not be treated as a strictly convex area because there could be multiple answers with different relaxation time or anticipation coefficients values. Due to non-convexity of the area, an optimization or search algorithm should be selected such that avoids stopping at local minima. Also, it is important to select an optimization or search algorithm which returns a set of answers, rather than one answer, to choose more meaningful parameters. To address these issues,

this study examined different search directions and found that if relaxation time, ϑ, is fixed and the search algorithm moves along the ϑ direction, trapping in local minima can be avoided. In addition, if the search algorithm starts moving from different relaxation time values, multiple answers can be found for calibration.

The proposed calibration methodology was applied for section lengths up to 450 ft and $\frac{L}{\Delta t}$ values between 50 mph and 90 mph. The results were used to develop a guideline to select the parameters of the second order model. The guideline recommends section length (L) of 450 ft be used to reduce computations. It suggests a range for $\frac{L}{\Delta t}$ based on the free flow speed, and a range for relaxation time (τ) based on both drivers' reaction time and the section length. The anticipation coefficient (ϑ) can be determined as a function of relaxation time using linear regression models. The linear models showed an increasing trend between relaxation time and anticipation coefficient, and slope and intercept values changed based on the values of L and $\frac{L}{\Delta t}$.

A modification was proposed to the second order model because the second order model returned slower queue propagation pattern and faster queue dissipation pattern than field data although it returned acceptable average queue length. The modification considered two different anticipation coefficients for queue propagation (ϑ_p) and queue shrinkage (ϑ_s). The two coefficients were calibrated for section length of 450 ft and $\frac{L}{\Delta t}$ of 80 mph and results showed that the ratio of $\frac{\vartheta_s}{\vartheta_p}$ ranges from 1.86 to 2.63. The reason that the ratios are greater than one is related to the fact that in the second order model, the acceleration rate is proportional to $\frac{\vartheta_s}{\text{Density in queue}}$ for vehicles departing queue and deceleration rate is proportional to $\frac{\vartheta_p}{\text{Density for arriving vehciles}}$ for vehicles joining queue. Density in the queue is larger than the density for arriving vehicles, thus ϑ_s has to be larger than ϑ_p to compensate a higher denominator.

The second order model was compared with the first order model and it was found that the first order model overestimated the queue length when the same speed-density relationships were used for both models. This happened because the second order model have calibration parameters of relaxation time and anticipation coefficient to improve the estimates, but the first order model did not have any parameter to calibrate. In order to improve queue lengths estimated by the first order model, several modifications were proposed. The modifications were:

Case 1) developing new regression function for congested condition, Case 2) new speed thresholds to declare queuing conditions, Case 3) increase free flow speed, and Case 4) decrease jam density.

First order model in the Case 3 returned better results than other cases. In particular, it underestimated average queue length by 177 ft which is corresponding to 8.6% error, and it overestimated maximum queue length by 1187 ft which is corresponding to 26.3% error. The standard deviation of error in queue length estimates was 955 ft which is corresponding to coefficient of variation of 46.5%.

The second order model returned better average and maximum queue lengths than the first order model, but a larger coefficient of variation. The discrepancy between the second order model estimates and the field data depended on the values of τ and ϑ. For example for τ =11 and ϑ =1.80, the second order model underestimated average queue length by 100.5 ft which is corresponding to 4.9% error. The maximum queue length was overestimated by 309 ft which is corresponding to 6.8% error. The standard deviation of error was 1734.7 ft which is corresponding to coefficient of variation of 84.4%.

The Two- ϑ model retuned better average queue length, maximum queue length, and coefficients of variation than the first order model and the second order model. The amount of discrepancy between estimates and field data depended on parameter values of τ, ϑ_p, and ϑ_s. For example for τ=11, ϑ_p=6.5, ϑ_s=13, average error in the queue length was 32 ft which is corresponding to 1.56% error. The maximum queue length was overestimated by 318 ft which is corresponding to 7.1% error. The standard deviation of error is 614.3 ft which is corresponding to coefficient of variation of 29.9%.

Thus based on the numerical comparisons, it can be concluded that the two- ϑ model is better than the first order model and the second order model. One may use the two- ϑ model when the rate of queue shrinkage and dissipation are significantly different. Consequently, one may use the second order model when the rates of dissipation and shrinkage are similar. In addition, in optimization programs using the second order model can significantly reduce computational complexity compared with the two- ϑ model which practically has a discontinuous function for ϑ. Thus this study uses the second order model to reduce computational complexities in the optimization program.

7.2 Conclusions on Work Zone Speed Harmonizer (WZSH)

This study developed and solved WZSH which is a mathematical program that simultaneously determines optimal advisory speeds, location of critical Changeable Message Signs (CMS), and number of critical CMSs for speed harmonization. The program minimizes an objective that is composed of total travel time and a penalty function for the number of critical CMSs.

The WZSH is a Mixed Integer Nonlinear Program (MINLP) and the main challenge to solve the problem is the presence of integer variables. According to source of integrality, solution methodologies were proposed:

- A continuous transformation approach was proposed to eliminate binary variables due to inclusion of piecewise speed-density relationships in the optimization program. Thus instead of solving a mixed integer nonlinear program, which is very challenging to solve for a large scale problem, one can solve the corresponding nonlinear program and significantly reduce computational difficulties. In addition, the transformation methodology provides the opportunity to use more realistic speed-density relationships, rather than very simplified models, and potentially enhances the reliability of speed harmonization techniques for field implementation. The methodology relies on some mathematical properties of the sub-functions, and it does not depend on the shape of the objective function or other constraints; as a result, one may apply the methodology for both minimization and maximization problems with different objective functions. The methodology was used to solve speed harmonization problem for a 12.5-mile roadway and a 50-min analysis duration where locations of the critical CMSs were given. For this problem, the methodology eliminated 30,000 binary variables and solved the problem in 23 minutes when the program includes roughly 158,000 variables and 186,000 constraints. The results showed that the speed harmonization can reduce delay by 15.7% and maximum queue length by 37.5% compared to the no speed harmonization condition.
- Three solution methods were proposed to determine number and location of critical CMSs. The solution methods are: 1) Greedy algorithm, 2) Augmented-Cut and Branch (AC&B) method, and 3) Approximate Decomposition method. The three methods were compared in three aspects:

1- Objective function value: It was shown that the difference between objective functions in the three methods is less than 1%. This means that the three methods returned similar objective functions values.

2- Number of runs: One run is defined as the solution of WZSH when there are no integer decision variables. This happens when the integer variables are relaxed or when values of zero or one are assigned to integer variables before problem solution. The number of runs is important because it indicates the required computational resources to solve the problem. For the benchmark problem, the greedy algorithm required 170 runs. Compared with the greedy algorithm, the AC&B method reduced the number of runs by a factor of 2.1, and the Approximate Decomposition method reduced the number of runs by a factor of 14.1. The AC&B method always required fewer number of runs than the greedy algorithm. In general, the Approximate Decomposition method required fewer number of runs than AC&B method. However, when the number of sections is small (e.g. less than 8), the AC&B method may require fewer runs; however this cannot be predicted before solving the problem because the number of runs in the AC&B method depends on the number of optimal critical CMSs and that is not known before solving the problem.

3- Type of objective function: The Approximate Decomposition method was developed for travel time minimization. For AC&B method and Greedy Algorithm, there is no assumption about the type of objective function. Thus these methods may be applied for any types of objective functions. As a result when the purpose of speed harmonization is to minimize travel time, it is recommended Approximate Decomposition method be used because it would require fewer number of runs. For other purposes, AC&B Method may be used.

7.3 Recommendations

This section first presents recommendations for similar applications based on the lesson learned from the conclusions. Then recommendations to expand the study are presented.

For similar applications the following recommendations are made:

- In order to calibrate the second order model, the study recommends a directional search to avoid stopping at local minima. For the search direction, this study recommends that relaxation time be fixed and the search algorithm move over the anticipation coefficient coordinate to avoid stopping at local minima.

- The study recommends excluding the portions of τ - ϑ space that overestimate or underestimate speed. These portions can be found based on the value of $\frac{\tau}{\vartheta}$. For starting values, one may use $\frac{\tau}{\vartheta}$=4 below which speed may be underestimated and $\frac{\tau}{\vartheta}$=10 above which speed may be overestimated.
- In general, the Two- ϑ model could be more accurate than the second order model; however the study recommends the second order model be considered to use when the rates of dissipation and shrinkage are similar since calibration and using the Two- ϑ model may involve computational complexities.
- For optimized speed harmonization, the study recommends using multi-regime traffic stream models which are more accurate than single-regime models in estimating congested traffic states. For this purpose, the proposed continuous transformation method can be used for continuous and piecewise traffic stream models to eliminate integer variables from the optimization program and consequently expedite the solution process.
- To determine location and number of critical CMSs, the study recommends using Approximate Decomposition method when the objective of the problem is to minimize total travel time because it would require less computational resources than other methods. For other types of objective functions, the study recommends using AC&B method.

For each item below, an issue and then recommended expansions to address the issue are discussed.

- To calibrate the parameters of the second order model, heavy vehicles in the field data were converted to passenger cars using a passenger car equivalency (PCE) factor of 1.5 proposed by Highway Capacity Manual 2010 for basic freeway section; thus computations assumed that traffic is composed of only passenger cars. It is recommended that field data be collected for different heavy vehicle percentages to study effects of different traffic composition on the PCE value.
- The guideline, which was developed to determine the parameters of the second order model, may be used for section lengths around 450 ft or shorter, but for longer section lengths (e.g. 0.5 miles), appropriate field data should be collected to expand the

guideline. In addition, the study recommends developing a similar guideline for the modified second order model in which two anticipation coefficients are used.

- The nonlinear transformation method was developed only for continuous and multi-regime models. New methodologies should be developed to include discontinues traffic stream models in the optimization program.
- This study assumes that a given CMS shows the same advisory speed for vehicles on different lanes. Alternatively, different advisory speeds may be assigned to vehicles on different lanes. To explore effectiveness of this alternative, the formulation should be expanded to i) consider effect of travel lane on traffic state, and ii) determine advisory speed for each travel lane.

References

1) Adeli, H., and X. Jiang, “Neuro-fuzzy logic model for freeway work zone capacity estimation”, Journal of Transportation Engineering, 2003.
2) Al-Kaisy, A., and F. Hall, “Effect of darkness on the capacity of long-term freeway reconstruction zones,” Transportation Research Circular, 2000.
3) Al-Kaisy, A. and F. Hall, “Examination of the effect of driver population at freeway reconstruction zones”, Transportation Research Record 1776, 2001.
4) Al-Kaisy, A., and F. Hall, “Guidelines for estimating freeway capacity at long-termre construction zones,” Journal of Transportation Engineering, Vol. 129, 2003.
5) Al-Kaisy, A., M. Zhou, and F. Hall, “New insights into freeway capacity at work zones: an empirical case study”, Transportation Research Record 1710, 2000.
6) Alessandri A., Di Febbraro A., Ferrara A. and Punta E., "Nonlinear optimization for freeway control using variable-speed signaling[J].," Vehicular Technology IEEE Transactions on, vol. 48(6), pp. 2042 - 2052, 1999.
7) Allaby, P.; Hellinga, B.; and Bullock, M., Variable speed limits: safety and operational impacts of a candidate control strategy for freeway applications. IEEE transaction 2007
8) American Highway Users Alliance, 1999-2004, Unclogging America’s arteries: effective relief for highway bottlenecks, http://www.highways.org/pdfs/ bottleneck2004.pdf.
9) Aw, A.; Rascle, M. Resurrection of second order models of traffic flow. SIAM journal of applied mathematics, 2000, 60 (3), pp. 916–938
10) Benekohal, R. F., A-Z Kaja-Mohideen, and M. V. Chitturi, “Methodology for estimating operating speed and capacity in work zones”, Transportation Research Record 1883, 2004.
11) Benekohal, R. F.; Ramezani, H.; Avrenli, K.; Delay and user’s cost in highway work zones, Research Report ICT-10-07. http://ict.illinois.edu/publications/report%20files/FHWA-ICT-10-075.pdf , 2010
12) Bertini R., Boice S. and Bogenberger K., "Dynamics of Variable Speed Limit System Surrounding Bottleneck on German Autobahn[J].," Transportation Research Record Journal of the Transportation Research Board, vol. 1978(1), pp. 149-159, 2006.
13) Berton, P.; Hegyi, A.; De Schutter,B.; Hellendoorn, H. , Shockwave elimination/reduction by optimal coordination of variable speed limit, IEEE proceeding 2002

14) Bham G. H.; Long S.; Baik H.; Ryan T.; and Gentry L., "Evaluation of Variable Speed Limits on I-270/I-255 in St. Louis," OR11-014, MoDOT, 2010.
15) Carlson R. C., Papamichail, I.; Papageorgiou, M; Local feedback-based mainstream traffic flow control on motorways using variable speed limits,, IEEE transactions on intelligent transportation systems, vol. 12, no. 4, December 2011, 1261-127
16) Carlson, R. C.; Papamichail, I.; Papageorgiou, M., Messmer, A., Optimal motorway traffic flow control involving variable speed limits and ramp metering, vol. 44, no. 2, 2010, pp. 238–253 transportation science,
17) Chang G-L., Park S. Y. and Paracha J., "Intelligent Transportation System Field Demonstration: Integration of Variable Speed Limit Control and Travel Time Estimation for a Recurrently Congested Highway[J].," Transportation Research Record: Journal of the Transportation Research Board, vol. 14(2243), pp. 55-66, 2011.
18) Chien, S.I.J., D.G. Goulias, S. Yahalom, S.M. Chowdhury, Simulation-based estimates of delays at freeway work zones. Journal of Advanced Transportation, 36 (2), 2002, 131-156.
19) Chitturi, M. C., and R. F. Benekohal., Comparison of QUEWZ, FRESIM and QuickZone with Field Data for Work Zones. CD-ROM, Transportation Research Board of the National Academies, Washington D.C. 2004
20) Chitturi, M. V.;Benekohal, R. F.; Kaja-Mohideen, A., Methodology for Computing Delays and Users Costs in Work Zones, In Transportation Research Record: Journal of the Transportation Research Board, No 2055, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 31-38.
21) Chitturi, M.V., R.F. Benekohal. Work zone queue length and delay methodology. CD-ROM, Transportation Research Board of the National Academies, Washington D.C 2009.
22) Collura, J., K. P. Heaslip, K. Moriarty, F. Wu, R. Khanta, and A. Berthaume., Simulation models for assessment of the impacts of strategies for highway work zones: eight case studies along interstate highways and state routes in New England. In Transportation Research Record: Journal of the Transportation Research Board, No 2169, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 62–69.
23) Cremer, M. der verkehrsfluß auf schnellstrassen. Springer verlag, Berlin, Germany, 1979.
24) Cremer, M.; M. Papageorgiou, Parameter identification for a traffic flow model , automatic vol. 17 No. 6 pp 837-843, 1981

25) Czyzyk, J.; Mesnier, M. P.; and Moré, J. J. The NEOS server. IEEE journal on computational science and engineering 5(3), 1998, pp. 68-75.
26) Daganzo, C.F.; The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory, Transportation. Research Part.-B. vol. 28b, no. 4, pp. 269-287, 1994.
27) Daganzo, C. F.; "A finite difference approximation of the Kinematic wave model of traffic flow", Transportation Research Part B, Vol 29B, pp.261-276, 1995.
28) Daganzo, C. F.; "Requiem for second-order fluid approximations of traffic flow", Transportation Research Part B, Vol 29B, pp.277-286, 1995.
29) Del castillo, J. M. ; P. Pintado and F. G. Benitez , The reaction time of drivers and the stability of traffic flow, transportation research part B. Vol. 28b, No. 1, pp 35-60, 1994
30) Dick, A. C., Speed/flow relationships within an urban area, traffic engineering control, 8, (6): 393-396 (1966)
31) Dixon, K. K., J. E. Hummer, and A. R. Lorscheider, "Capacity for north Carolina freeway work zones", Transportation Research Record 1529, 1996.
32) Dolan, E. The NEOS server 4.0 administrative guide. Technical memorandum anl/mcs-tm-250, mathematics and computer science division, Argonne national laboratory. 2001.
33) Edie, L. C. Car following and steady state theory for noncongested traffic, Oper. Res. , 9 (1): 66-76 (1961)
34) Federal Highway Administration, ATDM program brief: an introduction to active transportation and demand management, http://www.ops.fhwa.dot.gov/publications/fhwahop12032/, 2012.
35) Federal Highway Administration, Final rule on work zone safety and mobility. Federal Register/ Rules and Regulations, 2004, 69 (174).
36) Federal Highway Administration, Manual on Uniform Traffic Control Devices, Sec. 6c.04, 2009
37) Federal Highway Administration, Revised Monograph on Traffic Flow Theory, http://www.fhwa.dot.gov/publications/research/operations/tft/, 2012.
38) Garber, N. J., and S. Srinivasan, "Final report: effectiveness of changeable message signs in controlling vehicle speeds in work zones, phase ii",VTRC 98-R10, Virginia Transportation Research Council, December 1998.

39) Ghods, A.; Rahimi Kian, A.; Tabibi, M.; A genetic-fuzzy control application to ramp metering and variable speed limit control, IEEE proceeding 2007
40) Ghods, A.; Rahimi-Kian, A. A game theory approach to optimal coordinated ramp metering and variable speed limit. IEEE proceeding 2008
41) Grewal, M. S.; H.J. Payne , Identification of Parameters in a freeway traffic model, IEEE transactions on systems, man, and cybernetics, vol. smc-6, no. 3, march 1976
42) Hadiuzzaman, Md.; Qiu, T.; Cell transmission model based variable speed limit control for freeway. TRB 2012
43) Hegyi, A.; De schutter, B.; and Hellendoorn,J.; Optimal coordination of variable speed limits to suppress shock waves, IEEE transactions on intelligent transportation systems, vol. 6, no. 1, march 2005, 102-112
44) Highway Agency, "M25 Controller Motorways, Summary Report," 2007.
45) Jiang, X., H. Adeli, Freeway work zone traffic delay and cost optimization model, Journal of Transportation Engineering, 129 (3), 2003, 230-241.
46) Jiang, Y, A model for estimating excess user costs at highway work zones, In Transportation Research Record: Journal of the Transportation Research Board, No 1657, Transportation Research Board of the National Academies, Washington, D.C., 1999, pp. 31-41.
47) Jiang, Y., "Traffic capacity, speed, and queue-discharge rate of Indiana's four-lane freeway work zones", Transportation Research Record 1657, 1999
48) Johansson, G., and Rumar, H., Drivers' Reaction times, Human factor, 13 (1) 23:27, 1971.
49) Kamel, B.; Benasser, A.; Jolly D.; Flatness based control of traffic flow for coordination of ramp metering and variable speed limits. IEEE proceeding, 2008.
50) Kang, K.; Chang, G.; and Zou, N.; Optimal dynamic speed-limit control for highway work zone operations. TRR 2004
51) Karim, A., and H. Adeli, "Radial basis function neural network for work zone capacity and queue estimation", Journal of Transportation Engineering, 2003.
52) Kianfar, J.; Edara, P.; Sun, C. Operational analysis of a freeway variable speed limit system - case study of deployment in Missouri,
53) Kim, T., D. J. Lovell, and J. Paracha, "A new methodology to estimate capacity for freeway work zones", Transportation Research Board 80th Annual Meeting, 2001.

54) Kotsialos, A.; M. Papageorgiou, C. Diakaki, Y. Pavlis, and F. Middelham, Traffic flow modeling of large sclae motorway network using the maroscopic modeling tool METANET, IEEE transactions on intelligent transportation systems, vol. 3, No. 4, 2002, pp. 282-292
55) Kwon, E.; Brannan, D.; Shouman, K.; Isackson, C.; and Arseneau. B.; Development and field evaluation of variable advisory speed limit system for work zones. TRR 2007
56) Kwon, E.; Park, C.; Lau, D.; Kary, B.; Minnesota variable speed limit system: adaptive mitigation of shock waves for safety and efficiency of traffic flows. TRB 2012
57) Lee, C., D. A. Noyce, and X. Qin. , Development of traffic delay assessment tool for short-term closures on urban freeways. In Transportation Research Record: Journal of the Transportation Research Board, No 2055, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 39–48.
58) Lighthill, M.J.; G.B.Whitham, on kinematic waves II: a theory of traffic flow on long, crowded roads. proceedings of the royal society of London ser. a 229, 317-345, 1955
59) Lin, P.; Kang, K,; Chang, G.; Exploring the effectiveness of variable speed limit controls on highway work-zone operations. Intelligent Transportation Systems, 8:1–14, 2004
60) Lu, X.; Qiu, t.; Varaiya P.; Horowitz, R.; Shladover, S.; Combining variable speed limits with ramp metering for freeway. Traffic Control American Control Conference, 2010
61) Lu, X.; Varaiya, P.; Horowitz, R.; Su, D.; Shladover, S.; A novel freeway traffic control with variable speed limit and coordinated ramp metering. TRB 2011.
62) Maze, T.H.; Schrock, S. D.; Kamyab, A.; “Capacity of freeway work zone lane closures”, Mid-Continent Transportation Symposium 2000 Proceedings.
63) Macdonald, M. ATM monitoring and evaluation 4-lane variable mandatory speed limits 12 month report. Highways Agency, Bristol UK, 2008.
64) McMurtry, T.; Saito, M.; Riffkin, M.; Heath, S.; Variable speed limits signs: effects on speed and speed variation in work zones. TRB 2008.
65) Papageorgiou, M.; Some remarks on macroscopic traffic flow modelling, transportation research Part A, Vol. 32, No. 5, pp. 323-329, 1998
66) Papageorgiou, M.; Blosseville, J.; Hadj-Salem, H.; Macroscopic modeling of traffic flow on the boulevard Peripherique in Paris, Transportation Research Part B. pp. 29-47, 1989.

67) Papageorgiou. M.; J. Blosseville, H. Hadj-Salem , Modeling and real time control of traffic flow on the southern part of Boulevard Peripherique in Paris: Part I: Modeling, , Transportation Research Part A, vol. 24A, pp 345-359, 1990

68) Papageorgiou, M.; Kosmatopoulos,E.; Papamichail, I; Effects of variable speed limits on motorway traffic flow, transportation research record: journal of the transportation research board, no. 2047, transportation research board of the national academies, Washington, D.C., 2008, pp. 37–48.

69) Papageorgiou, M.; Papamichail, I.; Spiliopoulou, A.D.; Lentzakis, A.F. ; Real-time merging traffic control with applications to toll plaza and work zone management, transportation research part c 16 (2008) 535–553

70) Park, B.; Yadlapati, S.; Development and testing of variable speed limit logics at work zones using simulation, TRB 2002.

71) Payne, H. J., Models of freeway traffic and control, Simulation councils proceedings series, vol. 1, pp. 51-60, 1971

72) Piao, J.; and McDonald, M.; Safety impacts of variable speed limits – a simulation study. IEEE proceeding 2008

73) Ramezani, H.; Benekohal, R. F.; Avrenli, K.; Methodology to estimate moving queue length and delay in highway bottlenecks. Traffic Flow Theory conference proceeding. 2010.

74) Ramezani, H.; Benekohal, R. F.; Avrenli, K.; Methodology to measure capacity for work zones. Transportation Research Board. 2010.

75) Sarasua, W. A., W. J. Davis, D. B. Clarke, J. Kottapally, and P. Mulukutla, “Evaluation of interstate highway capacity for short-term work zone lane closures”, Transportation Research Record 1877, 2004.

76) Schnell, T., J. S. Mohror, and F. Aktan., Evaluation of traffic flow analysis tools applied to work zones based on flow data collected in the field. In Transportation Research Record: Journal of the Transportation Research Board, No 1811, Transportation Research Board of the National Academies, Washington, D.C. 2002, pp. 57–66.

77) Smulders, S. Control of freeway traffic flow by variable speed signs. Transportation research, vol. 24b, 1990, pp. 111–132.

78) Su, D.;Lu, X.;Varaiya, P.;Horowitz, R.;Shladover, S.; Variable speed limit and ramp metering design for congestion caused by weaving. TRB 2011

79) U.S. Department of Energy, May 2002, Temporary losses of highway capacity and impacts on performance, Oak Ridge National Laboratory (ORNL/TM-2002/3).

80) U.S. Department of Transportation, Federal Highway Administration, August 2002, A Snapshot of Peak Summer Work Zone Activity Reported on State Road Closure and Construction Websites, Washington, D.C.

81) U.S. Department of Transportation, 2009, Coordinating, planning and managing the effects of roadway construction with technology, Intelligent Transportation Systems Joint Program Office,
http://www.illinoistollway.com/pls/portal/url/ITEM/D17B2E300EBF4E01910D68BB5B61A4CE.

82) Vanderbei, R.J., An interior point code for quadratic programming, optimization methods and soft., 1999

83) Vanderbei, R.J.; Shanno, D.N.; An interior point algorithm for nonconvex nonlinear programming ,comput.Optim.App, 1999

84) Venugopal, S., and A. Tarko, "Investigation of factors affecting capacity at rural freeway work zones", Transportation Research Board 80th Annual Meeting, 2001.

85) Wang, Y.; Ioannou, P.; A new model for variable speed limits. TRB 2011

86) Yang X., Lu Y. and Chang G., "Proactive Optimal Variable Speed Limit Control for Recurrently Congested Freeway Bottlenecks [J].," Transportation Research Board Meeting, 2013.

87) Zackor, H.; Beurteilung verkehrsabhängiger geschwindigkeitsbeschränkungenauf autobahnen. strassenbau und strassenverkehrstechnik, vol. 128, 1972, pp. 1–61.

88) Zhang, H. Liu, J. Sun and D. Wang, "Variable Speed Limit Strategy to Improve the Safety and Environmental Impact of Freeway Traffic," in 94th Transportation Research Board Annual Meeting, Washington DC, 2015.

89) Zhang, H.M., A theory of nonequilibrium traffic flow. Transportation Research Part B 32, 1998, pp. 485–498

90) Zhang, H. M., "Comment on "On the controversy around Daganzo's requiem for and Aw-Rascle's resurrection of second-order traffic flow models" by D. Helbing and A.F. Johansson", The European Physical Journal B, pp563-568, 2009